

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2002, 09:27:40 ; Search time 22.69 seconds
(without alignments)
657.549 Million cell updates/sec

Title: US-09-882-434A-1

Perfect score: 551
Sequence: 1 MASTKLFPSVITVMMLTMA.....FGSSARACNPFCKSIFIOC 102

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0.8
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:
2: SP-bacteria:
3: SP-fungi:
4: SP-human:
5: SP-invertebrate:
6: SP-mammal:
7: SP-mhc:
8: SP-organella:
9: SP-phage:
10: SP-plant:
11: SP-rodent:
12: SP-virus:
13: SP-vertebrate:
14: SP-unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.5	14.1	486	3	005838 Q05838 saccharomyc
2	75.5	13.7	315	2	086833 086833 streptomyc
3	75.5	13.7	1358	5	09BUR5 09BUR5 leishmania
4	75	13.6	254	5	045025 045025 hydra magni
5	71	12.9	730	2	007667 007667 enterococcu
6	70	12.7	307	12	09WQ80 09WQ80 canine coro
7	70	12.7	307	12	09WQ79 09WQ79 feline infe
8	69.5	12.6	1792	13	057484 057484 gallus galli
9	69	12.5	1280	6	046605 046605 canis fami
10	68.5	12.4	125	4	000277 000277 homo sapien
11	68.5	12.4	372	4	000279 000279 homo sapien
12	67.5	12.3	184	5	009943 009943 caenorhabdi
13	67	12.2	307	12	09WQ75 09WQ75 porcine tra
14	67	12.2	653	12	09IBR0 09IBR0 spidoptera
15	67	12.2	1486	4	014637 014637 homo sapien
16	67	12.2	2121	12	09YMB7 09YMB7 porcine tra
17	66.5	12.1	6685	12	09IW06 09IW06 porcine tra
18	66.5	12.1	81	10	024225 024225 oryza sativ
19	66.5	12.1	251	2	092805 092805 streptomyc

20	66	12.0	534	3	074996 074996 yarrowia 11
21	66	12.0	1163	6	09RSU2 09RSU2 felis silve
22	65.5	11.9	197	3	09P7E1 09P7E1 schizosacch
23	65.5	11.9	233	5	045333 045333 caenorhabdi
24	65.5	11.9	244	10	09T0D2 09T0D2 arabidopsis
25	65.5	11.9	580	5	09U0B8 09U0B8 caenorhabdi
26	65.5	11.9	2346	11	09JLC1 09JLC1 mus musculi
27	65	11.8	199	5	094476 094476 dictyostell
28	65	11.8	1086	4	075808 075808 homo sapien
29	64.5	11.7	147	4	09H5U7 09H5U7 homo sapien
30	64	11.6	54	5	09H5U7 09H5U7 mesobuthus
31	64	11.6	767	13	09DGR2 09DGR2 xenopus lae
32	64	11.6	1272	11	09JRK4 09JRK4 rattus norv
33	64	11.6	2607	3	09Y8A2 09Y8A2 glibberella
34	64	11.6	2275	11	09WTS6 09WTS6 mus musculi
35	64	11.6	2824	13	09W7R3 09W7R3 brachydanio
36	64	11.6	3016	2	09W7R3 09W7R3 synchocyst
37	63.5	11.5	179	2	033801 033801 salmonella
38	63.5	11.5	213	10	09AML9 09AML9 oryza sativ
39	63.5	11.5	503	2	045834 045834 clostridium
40	63.5	11.5	511	2	094618 094618 clostridium
41	63.5	11.5	887	10	09LHCO 09LHCO arabidopsis
42	63.5	11.5	1083	5	026423 026423 carcinoscor
43	63.5	11.5	2093	5	09U131 09U131 leishmania
44	63	11.4	157	10	049178 049178 oryza sativ
45	63	11.4	1169	11	060502 060502 cricetus

ALIGNMENTS

RESULT 1
ID 005838 PRELIMINARY: PRT: 486 AA.
AC 005838:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HEXOKINASE PI (HXK2).
GN HXK2 OR YGL253W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi, Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86083199; PubMed=3908224;
RT Kopeckzki E., Entlian K.D., Mecke D.;
RT *Complete nucleotide sequence of the hexokinase PI gene (HXK1) of
RT Saccharomyces cerevisiae.*
RL Gene 39:95-101(1985).
DR EMBL: M14411; AAA34699.1; -.
DR HSP: P19367; HXK.
DR SGD: S0003222; HXK2.
DR InterPro: IPR001312; Hexokinase.
DR Pfam: PF00349; hexokinase.1.
DR PRINTS: PR00475; HEXOKINASE.
DR PRODOM: PD001109; Hexokinase.1.
DR PROSITE: PS00378; HEXOKINASES.1.
SQ SEQUENCE 486 AA; 53952 MW; 218F5AC9B8C36C8C CRC64;

Query Match 14.1% Score 77.5; DB 3; Length 486;
Best local similarity 31.2% Pred. No. 0.99; 27; Indels 31; Gaps 6;
Matches 30; Conservative 8; Mismatches 27; Indels 31; Gaps 6;

DB 8 FSVITVM--LIAMASVNGSAFTYWSGRCNNRRAERYSKGCSAIIHKGYSFYSYG 64
DB 369 FGINITVVERKLRIRLSLIGA-----RAARUSVCIAIICKRGK---TG 412
DB 65 OTAAALYNAGCGVAHTRF--GSSARACNP---FGW 95
DB 413 HIAA-----DGSVSTRPGFKKANAALMDIYGM 441

```

RESULT 2
ID 086833 PRELIMINARY: PRT: 315 AA.
AC 086833:
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DE PUTATIVE ALDEHYDE DEHYDROGENASE (FRAGMENT).
GN SC8B7.12C.
OS Streptomyces coelicolor.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1902;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parthill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RX MEDLINE=97000351; PubMed=8843436;
RA Klenzsch H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome."
DR EMBL: AL031225; CNA2223.1;
DR HSSP: P56533; 14AS.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd. 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 315 AA: 33701 MW: 42168DEBA0B9671 CRC64;

Query Match
Best Local Similarity 13.7%; Score 75.5; DB 2; Length 315;
Matches 30; Conservative 20; Mismatches 47; Indels 19; Gaps 5;

QY 1 MASTKLFSSVITVWML-----TAMSEWNGSAFTVWSGPCN-NRAERYSKCCSAIH 53
DB 199 VQSEIFGVLVLPEDTDEGRILANDPRLAASMSRDVYRANRATREIKAGCWIN 258
QY 54 Q-----KGGYDF$YGTOTALYNQAGCGVAHTRGSSARACNPFQW-KSLF 99
DB 259 DHIPISEMHGKYGKSGKGMASVSEFEYQVKNHVFDMTVAAK--DWHRTVF 312

RESULT 3
ID 09BLR5 PRELIMINARY: PRT: 1558 AA.
AC 09BLR5:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYPOHETICAL 165.8 KDA PROTEIN.
GN L65A.05.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_Taxid=5664;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Tosato V., Charlton L., Blanchettin G., Bruschi C.V., Ivens A.C.,

```

```

RA Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RT Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL512294; CAC29453.1;
KW Hypothetical protein.
SQ SEQUENCE 1558 AA: 165803 MW: BB029575FEDB8265 CRC64;

Query Match
Best Local Similarity 13.7%; Score 75.5; DB 5; Length 1558;
Matches 32; Conservative 10; Mismatches 33; Indels 55; Gaps 6;

QY 10 VTTVMILAMSEWNGSAFTVWSGPCNNRAERY-SKCGSAIHOKGYDFSTGOTAA 68
DB 663 VALLSWYAVDEAVFGRILRYW-----RLAQKHRSRCPAKSCGETGAD-----AAA 710
QY 69 LYN-----OAGCGVAH--TRGS-----S 86
DB 711 LNSDPWRGVVDAVRCYTDGDCGQAHGEPALSLGRCLLAIWLSLALVDYDGVGA 770
QY 87 ARACNPFQWK 96
DB 771 AAPCPMTWK 780

RESULT 4
ID 045025 PRELIMINARY: PRT: 254 AA.
AC 045025:
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE SPINMLIN.
OS Hydra magnipapillata (Hydra).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_Taxid=6085;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-SF-1; TISSUE=NEMATOCYST SPINES;
RA Koch A.W., Holstein T.W., Mala C., Kurz E., Engel J., David C.N.;
RL Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043907; AAC39121.1;
SQ SEQUENCE 254 AA: 25393 MW: 649BAB00B3BEABF9 CRC64;

Query Match
Best Local Similarity 13.6%; Score 75; DB 5; Length 254;
Matches 24; Conservative 14; Mismatches 24; Indels 26; Gaps 6;

QY 13 VMLIAMSEWNGSAFTVWSGPCNNRAERYSKCGSAIHOKGYDFSTGOTALYNG 72
DB 9 VLLVAVDAR-----PW-GPCADDSYGYGCG--HHQANGY-----GGAHNA 48
QY 73 AG-CGVAH-----TRGSSARACNPFQ 94
DB 49 AGCCNGLAHGHGAGVGOAHNHAGYGYG 76

RESULT 5
ID 007667 PRELIMINARY: PRT: 730 AA.
AC 007667:
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)

```


DR PROSITE: PS00022; EGF_1; UNKNOWN_10.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01248; LAMININ TYPE EGF; 12.
 KW EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.
 SQ SEQUENCE 1792 AA; 195722 MW; 4A4CB0206F6A600 CRC64;

Query Match 12.6%; Score 69.5; DB 13; Length 1792;
 Best Local Similarity 25.0%; Pred. No. 31;
 Matches 25; Conservative 9; Mismatches 41; Indels 25; Gaps 4;

OY 18 AMASEWNGAFYWSGPGCNNAERYK-----SKGCSAIIHQKGYD 59
 DB 300 ATADGVHGRVCVCKHHTQGLNCERCEDFYHELPRPAEGSSTNACRRCDCNEHSRCHFD 359

OY 60 ---FSYGTGATALYNQAGSGVAHTRFGSSARACNPGWK 96
 DB 360 MAVFLATGNTSG---AVCCGCCGHTNGRCHICKPYRK 395

RESULT 9
 ID 046605 PRELIMINARY; PRT: 1280 AA.

AC 046605:
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MULTIDRUG RESISTANCE P-GLYCOPROTEIN.

OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 -RN [1]

RP SEQUENCE FROM N.A.

RA Tissue-Colon:
 RA Puel O., LePage J.F., Alvine M., Galtier P., Plineau T.;
 Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

-I- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
 TRANSPORTERS).

CC EMBL: AF045016; AAC02113.1; -.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC transporter_tmam.
 DR InterPro: IPR003439; ABC transporter.
 DR InterPro: IPR001687; ATP-GTP_A.
 DR Pfam: PF00664; ABC_membrane_2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1280 AA; 141523 MW; 762DD5AF6AC73306 CRC64;

Query Match 12.5%; Score 69; DB 6; Length 1280;
 Best Local Similarity 25.5%; Pred. No. 25;
 Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVTVTMLAMASEWY-----NGSAFTWVGPGCNNAERYSKGCCSAIHQKGYD 59
 DB 335 VFFSVLIGAFSGISQASPSIEAFANRGAAVEIFKIIDNKPISIDYSKSHKPDNIKGNLE 394

OY 60 -----FSY-----TGGRALYNQAGC 75
 DB 395 FKNVHFSPSRKREKVLKGLNKLKVGSGTVALVQNSGC 432

RESULT 10
 ID 000277 PRELIMINARY; PRT: 126 AA.

AC 000277:
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE LYMPHOCYTE ASSOCIATED RECEPTOR OF DEATH 6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 -RN [1]

RP SEQUENCE FROM N.A.
 RA MEDLINE=97272273; PubMed=9114039;
 RA Scraton G.R., Xu X.N., Olsen A.L., Cowper A.E., Tan R.,
 RA McMichael A.J., Bell J.I.;
 RT "LAND: a new lymphoid-specific death domain containing receptor
 RT regulated by alternative pre-mRNA splicing."
 RT Proc. Natl. Acad. Sci. U.S.A. 94:4615-4619(1997).
 DR EMBL: U94507; AAC51312.1; .
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001368; TNFR_c6.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWN_1.
 DR SMART: SM00208; TNFR; 1.

SQ SEQUENCE 126 AA; 13349 MW; EBAC9BD7FB3EDB68 CRC64;

Query Match 12.4%; Score 68.5; DB 4; Length 126;
 Best Local Similarity 27.0%; Pred. No. 2.6;
 Matches 27; Conservative 12; Mismatches 26; Indels 35; Gaps 8;

OY 9 SVITVMILAMASEWNGSAFTWVGPGCNNAERYK-----CGCSAIIHQKGYDFSY 62
 DB 10 AVAALLLVLLGARAQGGT-----RSPRCDCAGDPHKRIGLFCRCGCPAASQ----- 56

OY 63 TGGTAAALYNQAGSGVAHTRFGSSARACNPGWKSIIFQC 102
 DB 57 -----VALEN-----CSAVADTRCG-----CKP-GW---FVQC 80

RESULT 11
 ID 000279 PRELIMINARY; PRT: 372 AA.

AC 000279:
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE LYMPHOCYTE ASSOCIATED RECEPTOR OF DEATH 8 (LYMPHOCYTE ASSOCIATED
 DE RECEPTOR OF DEATH 5).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 -RN [1]

RP SEQUENCE FROM N.A.
 RA Scraton G., Xu X.N., Olsen A., Cowper A., Tan R., McMichael A.,
 RA Bell J.I.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

CC EMBL: U94509; AAC51314.1; .
 DR EMBL: U94506; AAC51311.1; .
 DR HSSP: P19438; TEXT.
 DR InterPro: IPR000488; Death.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001368; TNFR_c6.
 DR Pfam: PF00531; death_1.
 DR SMART: SM00005; DEATH; 1.
 DR SMART: SM00208; TNFR; 1.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW Alternative splicing.
 SQ SEQUENCE 372 AA; 40391 MW; D4F87353AB7F40F6 CRC64;

Query Match 12.4%; Score 68.5; DB 4; Length 372;
 Best Local Similarity 27.0%; Pred. No. 7.9;
 Matches 27; Conservative 12; Mismatches 26; Indels 35; Gaps 8;

QY 9 SVITVMMIIMASEMVGSAFTWSPGCCNNRAERYSK-----C-GCSAIHOKGQYDFSY 62
Db 10 AVAAALLLVLLGARAAGCT-----RSPRCDCAGDFHKKIGLFCRCGPASQ-----56
QY 63 TGGTAALYNAGCGVAHTRFSSARACNPPGKMSITPQC 102
Db 57 -----VALEN---CSAVADTRCG-----CKP-GW---FVEEC 80

RESULT 12

QY 009943 PRELIMINARY: PRT: 184 AA.
AC 009943;
DT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
DE 01-FEB-1997 (TREMBLREL. 02, Last annotation update)
DE HYPOTHEMETICAL 20.1 KDA PROTEIN FL2A10.1 IN CHROMOSOME II.
GN FL2A10.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Gelael C.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U28731; AAA68294.1;
DR WormPep: FL2A10.1; CE01904.
KW Hypothetical protein.
SQ SEQUENCE 184 AA: 20062 MW: 4F19A9A6C96B175A CRC64;

Query Match 12.3%; Score 67.5; DB 5; Length 184;
Best Local Similarity 24.1%; Pred. No. 5;
Matches 26; Conservative 11; Mismatches 34; Indels 37; Gaps 4;

QY 1 MASTKLFESYITVMMIIMASEMVG-----GSAFTWAG-----PG 36
Db 78 MHSTTLFEVVEGLVAVSLAFDNLNPKNLQWRSSGSRVKGWGGGCGYGGCGYFG 137
QY 37 CNNRAERYSKCGCSAIHOKGQYDFSYTGTAALYNAGCGVAHTRF 84
Db 138 -----GYSGYG-----GGYPMWYGGCGMGSGYSSSMGYSRMSG 172

RESULT 13

QY 09W075 PRELIMINARY: PRT: 307 AA.
AC 09W075;
DT 01-NOV-1999 (TREMBLREL. 12, Created)
DT 01-NOV-1999 (TREMBLREL. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLREL. 12, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (FRAGMENT).
GN POL.
OS Porcine transmissible gastroenteritis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11149;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE-99319897; PubMed-10392726;
RA Stephensen C.B., Casebolt D.B., Gangopadhyay N.N.;
RT *Phylogenetic analysis of a highly conserved region of the polymerase
RT gene from 11 coronaviruses and development of a consensus polymerase
RT chain reaction assay."
RL Virus Res. 60:181-189(1999).
DR EMBL: AF124992; AAD32996.1;
KW RNA-directed RNA polymerase;
FT NON_TER 1
FT 307 307
SQ SEQUENCE 307 AA: 34500 MW: B3215EE59E9A21EF CRC64;

Query Match 12.2%; Score 67; DB 12; Length 307;
Best Local Similarity 25.0%; Pred. No. 9.6;
Matches 26; Conservative 12; Mismatches 36; Indels 30; Gaps 4;

QY 16 LIMASEMVGSAFTWSPGCCNNRAERYSKG-----CSAIHOKGQYDFSYGQTAA 68
Db 124 MIRMASMILGSKHV-----GCCTHNDREYRLSNELAQVLTVEVHCTGGFYFKPGGTTSG 178
QY 69 LYNAGCGVAHTRF-----GSSARACNPPGKMSI 98
Db 179 -----DGTAVANSAENIFQAVSANVKLLGVDNSNACNNVTAKSI 218

RESULT 14

QY 091BRO PRELIMINARY: PRT: 653 AA.
AC 091BRO;
DT 01-OCT-2000 (TREMBLREL. 15, Created)
DT 01-OCT-2000 (TREMBLREL. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLREL. 15, Last annotation update)
DE ORF131 p74.
OS Spodoptera exigua nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=10454;
RN [1]
RP SEQUENCE OF 596-653 FROM N.A.
RX MEDLINE-93286555; PubMed-8509757;
RA Zuidema D., van Oers M.M., van Strien E.A., Caballero P.C., Klok E.J.,
RA Zuidema D., van Oers M.M., van Strien E.A., Broer R., Zuidema D.,
RA Goldbach R.W., Vlak J.M.;
RT *Nucleotide sequence and transcriptional analysis of the p10 gene of
RT Spodoptera exigua nuclear polyhedrosis virus."
RL J. Gen. Virol. 74:1017-1024(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20036646; PubMed-10567663;
RA Zuidema D., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
RA Goldbach R.W., Vlak J.M.;
RT *Sequence and organization of the Spodoptera exigua multicapsid
RT nucleopolyhedrovirus genome."
RL J. Gen. Virol. 80:3289-3304(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Zuidema D.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RL Zuidema D., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
RA Goldbach R.W., Vlak J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF169823; AAF33660.1;
SQ SEQUENCE 653 AA: 74213 MW: DADAD68F070FE0BC CRC64;

Query Match 12.2%; Score 67; DB 12; Length 653;
Best Local Similarity 33.3%; Pred. No. 21;
Matches 19; Conservative 9; Mismatches 15; Indels 14; Gaps 3;

QY 43 RYKCGCSAI-----HOKGYD-----FSYTGTAALYNAGC---SGVAHTRF 85
Db 65 KFSKRGCSMTCYPHERGPIDANTPANYTOTSETAILYADPACYNLDRAVATREGA 121

RESULT 15

QY 014637 PRELIMINARY: PRT: 1486 AA.
AC 014637;
DT 01-JAN-1998 (TREMBLREL. 05, Created)
DT 01-JAN-1998 (TREMBLREL. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
DE LAMININ ALPHA 3B CHAIN (FRAGMENT).
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98055454; PubMed=9395076;
 RA Dollana R., Bellina I., Bucciolotti F., Mongiat M., Perris R.,
 RA Colombatti A.;
 RT "The human alpha3b is a 'full-sized' laminin chain variant with a more
 RT widespread tissue expression than the truncated alpha3a.";
 RL FEBS Lett. 417:65-70(1997).
 DR EMBL: AF005258; AAC51867.1.;
 DR HSSP; P02468; TITLE: EGF-like;
 DR InterPro: IPR000561; EGF-like;
 DR InterPro: IPR002049; Laminin_EGF;
 DR InterPro: IPR01886; LAMNT.
 DR Pfam: PF00053; Laminin_EGF; 10.
 DR Pfam: PF00055; Laminin_Nterm; 1.
 DR PRINTS: PR00011; EGFLAMININ.
 DR PRODOM: PD002082; LAMNT; 1.
 DR SMART; SM00180; EGF_Lam; 10.
 DR SMART; SM00001; EGF_Like; 1.
 DR SMART; SM00136; LAMNT; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_10.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 9.
 KW Glycoprotein; Laminin EGF-like domain; Repeat.
 FT NON_TER 1
 FT NON_TER 1486
 SQ SEQUENCE 1486 AA; 162495 MW; F2E5E87EB522E98C CRC64;

Query Match 12.2%; Score 67; DB 4; Length 1486;
 Best Local Similarity 28.1%; Pred. NO. 49;
 Matches 27; Conservability 8; Mismatches 23; Indels 38; Gaps 7;
 QY 34 GPGCUN-RAERYS-----KCSAITHOKGYDF---SYTQ-----TAAL 69
 DB 480 GPRCDICRGCFSPICQCMCSAL---GSYMPCCSVTGCECRPGVTGQRCDRLSGA 536
 QY 70 YNAGCGGVATHTFGSSARACNPFG---WKSIFIQ 102
 DB 537 YDFPHCG-----SSSACDPAGTINWNLGYCOC 564

Search completed: January 24, 2002, 09:29:32
 Job time: 112 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2002, 09:24:50 ; Search time 23.96 seconds

(without alignments)
315.337 Million cell updates/sec

Title: US-09-882-434a-1
Perfect score: 531
Sequence: 1 MASTKLFPSVTVMMLAMA.....FGSSARACNPFGRKSIPIQC 102

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSB/gcgdata/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	551	100.0	102	AAW31738	Macadamia integrifolia
2	74.5	13.5	486	AAAB10456	S. pastorianus mut
3	71	12.9	1280	AAAB1064	Cynomolgus monkey
4	71	12.9	1283	AAAB1065	Cynomolgus monkey
5	70.5	12.8	524	AAW34998	Archaeobacterium AE
6	69	12.5	1280	AAAB1068	Dog P-glycoprotein
7	69	12.5	1280	AAE00304	Dog P-glycoprotein
8	69	12.5	1281	AAE00303	Dog (Pgp) P-glycop
9	69	12.5	1281	AAE00308	Dog P-glycoprotein
10	69	12.5	1281	AAE00309	Dog P-glycoprotein
11	69	12.5	1281	AAE00310	Dog P-glycoprotein

12	68.5	12.4	1280	19	AAW48998	N-terminal single
13	68.5	12.4	1280	19	AAW49000	Human P glycoprote
14	68	12.3	1280	8	AAW70452	Sequence encoded b
15	68	12.3	1280	11	AAW04868	Protein encoded by
16	68	12.3	1280	14	AAW44297	Sequence encoded b
17	68	12.3	1280	15	AAW63624	Human P-glycoprote
18	68	12.3	1280	18	AAW4073	Human multilidug re
19	68	12.3	1280	19	AAW48997	wild-type human p
20	68	12.3	1280	19	AAW48999	C-terminal single
21	68	12.3	1280	21	AAW58186	Human wild-type mu
22	68	12.3	1280	21	AAW58187	Human G185v mutant
23	68	12.3	1280	22	AAW04347	Human multilidug re
24	68	12.3	1280	22	AAW40366	Human P-glycoprote
25	68	12.3	1280	22	AAW41959	Human MRL Homo
26	68	12.3	1280	22	AAE00306	Human P-glycoprote
27	67.5	12.3	902	18	AAW1285	H. armigera granul
28	67	12.2	1470	18	AAW23411	Porcine transmissi
29	66.5	12.1	151	21	AAW36342	Salmonella enterit
30	66.5	12.1	376	22	AAW42001	Human polypeptide
31	66.5	12.1	378	22	AAW40215	Human polypeptide
32	66	12.0	76	22	AAW4058	Plant defensin
33	65.5	11.9	244	21	AAW42513	Arabidopsis thalia
34	65.5	11.9	244	21	AAW44590	Arabidopsis thalia
35	65	11.8	223	21	AAW44260	Asparagus officina
36	64	11.6	1272	21	AAW70596	Rat multilidug resi
37	64	11.6	1272	21	AAW70597	Human laminin 2 ma
38	64	11.6	1576	21	AAW19802	Human laminin 8 po
39	64	11.6	1576	21	AAW48453	Human laminin 2 ga
40	64	11.6	1584	21	AAW19804	Human laminin 2 ga
41	64	11.6	1609	19	AAW50898	Human laminin 8 po
42	64	11.6	1609	21	AAW19801	Human laminin 2 ga
43	64	11.6	1609	21	AAW48452	Human laminin 8 po
44	64	11.6	1617	21	AAW19803	Human laminin 2 ga
45	63.5	11.5	126	21	AAW54746	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AAW31738	standard; Protein: 102 AA.
ID	AAW31738	
AC	AAW31738	
DE	27-MAR-1998	(first entry)
XX	Macadamia integrifolia antimicrobial protein 1.	
XX	Antimicrobial protein 1; growth inhibition; fungus; bacterium;	
KW	fungi; bacteria; pathogen control.	
OS	Macadamia integrifolia.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..26
FT		/label= sig-peptide
FT	Peptide	27..102
FT		/label= mat-peptide
XX		
PN	W09728185-A1.	
XX		
PD	07-AUG-1997.	
XX		
PF	31-JAN-1997;	97WO-AU00052.
XX		
PR	31-JAN-1996;	96AU-0007802.
XX		
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.	
XX		
PI	Goulter KC, Green JL, Harrison SJ, Manners JM, Marcus JP;	
XX		
DR	WPI, 1997-448317/41.	

DR N-PSDB: AAT88851.
 XX New antimicrobial protein from *Macadamia integrifolia* - active in
 PT inhibiting the growth of fungi and bacteria in plants and animals
 XX
 XX Claim 1; Pages 23-24; 38pp; English.
 CC The present sequence is *Macadamia integrifolia* antimicrobial
 CC protein 1, which exerts inhibitory activity on the growth of fungi
 CC and bacteria. It can be used for the control of pathogens in plants
 CC and animals.
 XX
 SO Sequence 102 AA;

Query Match 100.0%; Score 551; DB 18; Length 102;
 Best Local Similarity 100.0%; Pred. No. 7.8e-55;
 Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MASTKLFPSVITVMMILIMASEMNGSAFTWSPGCGNNRAERYSKCGSAIHQKGYDF 60
 DB 1 mastklfsvltvmmilimasemngsafvws9pgcnnraeryskcgcsahqkgydf 60
 OY 61 SYTGQTALYNOAGCSGVANTRFSSARACNPFQWKSIFIOC 102
 DB 61 sytgqtaalynagcsgvahrtrfssaracnfpqwkisfiicg 102

RESULT 2

AAB10456
 ID AAB10456 standard; Protein: 486 AA.

AC AAB10456;

DT 11-DEC-2000 (first entry);

DE 5. *pastorianus* mutant hexokinase protein.

KW Hexokinase; mutant; creatinine kinase detection; glucose detection.

OS *Saccharomyces pastorianus*.

PN JP2000175688-A.

PD 27-JUN-2000.

PE 17-DEC-1998; 98JP-0359018.

PR 17-DEC-1998; 98JP-0359018.

PS (TOYM) TOYOH K.

DR WPI: 2000-501189/45.

DR N-PSDB: AAA71483.

PT Mutant hexokinase, useful for the detection of creatinine kinase and
 PT glucose, comprises a deletion, addition or insertion in the wild-type
 PT amino acid sequence -

PS Claim 3; Page 8-9; 13pp; Japanese.

CC This invention describes a novel mutant hexokinase (I) comprising a
 CC protein having hexokinase activity in which at least one amino acid
 CC is added, deleted, inserted or replaced. The amino acid sequence
 CC constitutes the protein having hexokinase activity which remains after
 CC the protein is treated at 50 degrees C for 30 minutes in liquid state,
 CC and which is higher than that of the protein before it is mutated. (1)
 CC is used for the detection of creatinine kinase and glucose. This sequence
 CC represents the *Saccharomyces pastorianus* hexokinase protein which is
 CC described in the method of the invention.

SO Sequence 486 AA;

Query Match 13.5%; Score 74.5; DB 21; Length 486;
 Best Local Similarity 29.2%; Pred. No. 3;
 Matches 28; Conservative 10; Mismatches 27; Indels 31; Gaps 6;

OY 8 FSVITVMM--LIIMASEMNGSAFTWSPGCGNNRAERYSKCGSAIHQKGYDFSYTG 64
 DB 369 fgvitvgerklirrlisllga-----raarlvsglaavcqrkyk---lg 412
 OY 65 QTMALYNAGCSGVANTRF-GSSARACNPFQWKSIFIOC 95
 DB 413 hiaa-----dsvynrypyfkakaana[kdi]y 441

RESULT 3

AAB81064
 ID AAB81064 standard; Protein: 1280 AA.

AC AAB81064;

DT 25-JUN-2001 (first entry)

DE Cynomologous monkey P-glycoprotein variant 1.

KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRI;
 KW efflux pump.

OS *Macaca fascicularis*.

PN WO200123565-A1.

PD 05-APR-2001.

PE 28-SEP-2000; 2000MO-US26592.

PR 28-SEP-1999; 99US-0156821.

PR 12-OCT-1999; 99US-0158818.

PA (GENT-) GENTEST CORP.

PI Stocker PJ, Steimel-Crespi DT, Crespi CL;

DR WPI: 2001-316136/33.

DR N-PSDB: AAF86127.

PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
 PT (PGP) and homologous PGP polypeptides are useful for predicting
 PT bioavailability of compound and increasing PGP transporter activity in
 PT cell -

PS Claim 9; Page 57-59; 84pp; English.

CC This invention relates to a polynucleotide sequence encoding a
 CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
 CC PGP protein. PGP, also known as multidrug transporter, MDRI is a member
 CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
 CC exporting small molecules across the cell membrane. The invention
 CC includes a cynomologous monkey (*Macaca fascicularis*) PGP coding sequence
 CC and protein, and also that of an allelic variant. The PGP polynucleotide
 CC sequence is useful for increasing PGP transporter activity in a cell.
 CC Antisense sequences of the CDNA are useful for inhibiting PGP transport
 CC activity in a mammalian cell. They may also be used for increasing the
 CC bioavailability of a drug. The present sequence represents the
 CC cynomologous monkey P-glycoprotein.

SO Sequence 1280 AA;

Query Match 12.9%; Score 71; DB 22; Length 1280;
 Best Local Similarity 25.5%; Pred. No. 23;
 Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFPSVITVMMILIMASEMNGSAFTWSPGCGNNRAERYSKCGSAIHQKGYDF 59

Db 334 vffsvllgafsvqspgleafanargafefkldnkpdsidsysksghkpdnlkgnle 393
 QY 60 -----FSY-----TGOTALYNQAGC 75
 Db 394 frnvhtfyspsrkevklkgnlnkwsgqglvalvgnsgc 431

RESULT 4

AAB81065
 ID AAB81065 standard; Protein: 1283 AA.
 AC AAB81065;
 XX
 XX
 DT 25-JUN-2001 (first entry)
 DE Cynomologous monkey P-glycoprotein variant 2.
 XX
 XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
 KM efflux pump.
 KW
 XX
 OS Macaca fascicularis.
 XX
 FH Key Location/Qualifiers.
 FT Misc-difference 93..95
 FT "An additional 3 amino acids are present compared
 to PGP variant AAB81064"

MO200123565-A1.
 XX
 XX
 PD 05-APR-2001.
 PF 28-SEP-2000; 2000MO-US26592.
 XX
 XX
 PR 28-SEP-1999; 99US-0156921.
 PR 12-OCT-1999; 99US-0158818.
 XX
 PA (GENT-) GENTEST CORR.
 XX
 PI Stocker PJ, Steinml-Crespi DT, Crespi CL;
 DR WPI: 2001-316136/33.
 DR N-PSDB; AAF86128.
 XX
 XX
 PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
 PT (PGP) and homologous PGP polypeptides are useful for predicting
 PT bioavailability of compound and increasing Pgp transporter activity in
 PT cell
 XX
 PS Claim 9; Page 65-68; 84pp; English.

CC This invention relates to a polynucleotide sequence encoding a
 CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
 CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
 CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
 CC exporting small molecules across the cell membrane. The invention
 CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
 CC and protein, and also that of an allelic variant. The PGP polynucleotide
 CC sequence is useful for increasing PGP transporter activity in a cell.
 CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
 CC activity in a mammalian cell. They may also be used for increasing the
 CC bioavailability of a drug. The present sequence represents the
 CC cynomologous monkey P-glycoprotein variant 2. The protein has an
 CC additional 3 amino acids when compared to PGP variant 1 (AAB81065).

Sequence 1283 AA;

Query Match 12.9%; Score 71; DB 22; Length 1283;

Best Local Similarity 25.5%; Pred. No. 23;
 Matches 25; Conservative 12; Mismatches 33; Indels 26; Gaps 3;

OY 6 LFPSVITVWMLIAMSEMV-----NCSAFTVWSPGCCNNRAERYSKCGCSAITHRGYD 59

Db 337 vffsvllgafsvqspgleafanargafefkldnkpdsidsysksghkpdnlkgnle 396
 QY 60 -----FSY-----TGOTALYNQAGC 75
 Db 397 frnvhtfyspsrkevklkgnlnkwsgqglvalvgnsgc 434

RESULT 5

AAW34998
 ID AAW34998 standard; Protein: 524 AA.
 AC AAW34998;
 XX
 XX
 DT 21-MAY-1998 (first entry)
 DE Archaeobacterium AEP11a endoglucanase.
 XX
 XX
 KM Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
 KM biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
 KM thermostable enzyme; thermophilic; glycosidase.
 XX
 OS Archaeobacterium isolate AEP11a (Clone 63GP1).
 XX
 PN WO9744361-A1.
 PD 27-NOV-1997.
 PF 22-MAY-1997; 97WO-US08793.
 PR 22-MAY-1996; 96US-0651572.
 XX
 XX
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
 PI Lam DE, Mathur EJ;
 DR WPI: 1998-018435/02.
 DR N-PSDB; AAT94206.
 XX
 XX
 PT Endoglucanase(s), preferably form archaeal bacterium, AEP11a -
 PT useful to degrade carboxymethylcellulose and hydrolyse of
 PT beta-1,4-glycosidic bonds in cellulose
 XX
 PS Claim 1; Fig 1N; 164pp; English.

CC This protein comprises an endoglucanase of archaeobacterium AEP11a
 CC (Clone 63GP1), a hydrothermal vent isolate. The endoglucanase is
 CC capable of degrading carboxymethylcellulose and of hydrolysing the
 CC beta-1,4-glycosidic bonds in cellulose. It has homology to another
 CC endoglucanase (see AAW34985) of archaeobacterium AEP11a. It can be
 CC produced from native cells or from recombinant host cells,
 CC especially prokaryotic host cells transformed with a plasmid or
 CC virus-derived vector including the endoglucanase DNA (see AAT94206).
 CC 24 Endoglucanases (see AAW34986-W35008) are claimed. They can be
 CC used to degrade cellulose for the conversion of plant biomass into
 CC fuels and chemicals, for use in detergents, textiles, animal feed,
 CC waste treatment, and in the fruit juice and brewing industries for
 CC the clarification and extraction of juices.

Sequence 524 AA;

Query Match 12.8%; Score 70.5; DB 19; Length 524;

Best Local Similarity 26.4%; Pred. No. 9.3;
 Matches 33; Conservative 12; Mismatches 37; Indels 43; Gaps 8;

OY 3 STRLEFSV-ITVWMLIAM-----SEMVNGSAFT--VWSGPG-----CNR 40

Db 63 smkfkfalfilftfvmavqpasaaakyselsgylmqafydvpggslwdtlrsk 122

OY 41 AERYSKCGCSAL-----HOKGG-----YDESYTGOTALYNQAGCSGVATRRSS 86

Db 123 lpewyeaqaisalwlpakgmsgygmydydfldlge---ynq---kgtletrfysk 175

OY 87 ARACN 91
DB 176 gelin 180

RESULT 6

AAE0304
ID AAE0304 standard; Protein; 1280 AA.

AC AAE0304;

DT 13-JUN-2001 (first entry)

DE Dog P-glycoprotein (PGP) #2.

DE Dog P-glycoprotein SEQ ID 7.
KW Cynomolgus monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW efflux pump; dog.
OS Canis familiaris.

OS Canis familiaris.

PN WO200123565-A1.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26592.

PR 28-SEP-1999; 99US-0156921;
PR 12-OCT-1999; 99US-0158818.

PA (GENT-) GENTEST CORP.

PI Stocker PJ, Steimel-Crespi DT, Crespi CL;

PI WPI: 2001-316136/33.

DR WPI: 2001-316136/33.

PT Novel isolated nucleic acid encoding cynomolgus monkey P-glycoprotein

PT (PGP) and homologous PGP polypeptides are useful for predicting

PT bioavailability of compound and increasing PGP transporter activity in

PT cell -

PS Claim 8; Page 73-76; 84pp; English.

CC This invention relates to a polynucleotide sequence encoding a

CC cynomolgus monkey P-glycoprotein (PGP), and an allelic variant of the

CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member

CC of the ABC transporter superfamily. The enzyme serves as an efflux pump

CC exporting small molecules across the cell membrane. The invention

CC includes a cynomolgus monkey (Macaca fascicularis) PGP coding sequence

CC and protein, and also that of an allelic variant. The PGP polynucleotide

CC sequence is useful for increasing PGP transporter activity in a cell.

CC Antisense sequences of the cDNA are useful for inhibiting PGP transport

CC activity in a mammalian cell. They may also be used for increasing the

CC bioavailability of a drug. The present sequence represents

CC P-glycoprotein from a dog.

XX Sequence 1280 AA;

XX

Query Match 12.5%; Score 69; DB 22; Length 1280;

Best Local Similarity 25.5%; Pred. No. 38;

Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVITVWMLIMASENV-----NSAFYVWSGPGCNRNRAERYSKGCSAIIHQKGYD 59

DB 335 vffevllgafslgqaspsleatfanargaayelfklidnkpsidsysgshkpdnknle 394

OY 60 -----FSY-----TGQTAALYNQAGC 75

DB 395 fknvhtsyprkxevkllkgnlnkvsgqgtvalvngsc 432

RESULT 7

AAE0304
ID AAE0304 standard; Protein; 1280 AA.

AC AAE0304;

DT 13-JUN-2001 (first entry)

DE Dog P-glycoprotein (PGP) #2.

DE Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model.

OS Canis familiaris.

PN WO200123540-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26767.

PR 28-SEP-1999; 99US-0156510.

PA (GENT-) GENTEST CORP.

PI Stocker PJ, Steimel-Crespi DT, Crespi CL, Reif TC, Patten CJ;

PI WPI: 2001-235373/24.

DR N-PSDB; AAD03469.

PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog

PT PGP inhibitors -

PS Example 2; Page 72-75; 11pp; English.

CC The invention relates to dog P-glycoprotein (PGP) also referred

CC as multidrug transporter (MDR1) and nucleic acids encoding them.

CC The invention also includes fragments and biologically functional

CC variants of dog P-glycoprotein. PGP and their nucleic acids are

CC useful for determining the bioavailability of drugs and for

CC screening PGP inhibitors. They are useful for the diagnosis and

CC treatment of conditions characterised by PGP activity, by

CC reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic

CC acids are useful as antisense oligonucleotides, to induce a PGP

CC 'knockout' phenotype. They are used to prepare a non-human

CC transgenic animal, which are valuable as genetic models for

CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP). The

CC PGP enzyme functions as an efflux pump exporting small molecules

CC across the cell membrane. This enzyme is a member of the ABC

CC transporter family.

XX Sequence 1280 AA;

XX

Query Match 12.5%; Score 69; DB 22; Length 1280;

Best Local Similarity 25.5%; Pred. No. 38;

Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVITVWMLIMASENV-----NSAFYVWSGPGCNRNRAERYSKGCSAIIHQKGYD 59

DB 335 vffevllgafslgqaspsleatfanargaayelfklidnkpsidsysgshkpdnknle 394

OY 60 -----FSY-----TGQTAALYNQAGC 75

DB 395 fknvhtsyprkxevkllkgnlnkvsgqgtvalvngsc 432

RESULT 8

AAE0303
ID AAE0303 standard; Protein; 1281 AA.

AC AAE00303;
 XX 13-JUN-2001 (first entry)
 DT
 DE Dog (PGP) P-glycoprotein (genotype G) #1.
 XX
 DE Dog: P-glycoprotein; PGP: multidrug transporter; MDRL;
 KM drug bioavailability; transgenic animal; genetic model.
 XX
 OS Canis familiaris.
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 DR MPI: 2001-235373/24.
 DR N-PSDB; AAD03488.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors.
 XX
 PS Claim 17; Page 64-66; 11pp; English.
 XX
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDRL) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) also referred
 CC as genotype C protein. The PGP enzyme functions as an efflux
 CC pump exporting small molecules across the cell membrane. This
 CC enzyme is a member of the ABC transporter family.
 CC
 SO Sequence 1281 AA:
 Query Match 12.5%; Score 69; DB 22; Length 1281;
 Best Local Similarity 25.5%; Pred. No. 38;
 Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;
 QY 6 LFFSVITVMMILIMASEMV-----NGSAFTVWSGPCNNNAERYSKCGCSAITHQGYD 59
 DB 336 VFTSVLLGAFSGAASLEAFANARGAAEYELIKLIDNKPISYSKSHKPDNIKGNLE 395
 QY 60 -----FSY-----TGQTAALYNQAC 75
 DB 396 fknvhfsypsrkevklkglnlkvsgqvalvngsgc 433
 RESULT 9
 AAE00308
 ID AAE00308 standard; Protein: 1281 AA.
 XX
 AC AAE00308;
 XX
 DT 13-JUN-2001 (first entry)

XX
 DE Dog P-glycoprotein (PGP) allelic variant (Genotype A).
 XX
 DE Dog: P-glycoprotein allelic variant; PGP: multidrug transporter;
 KM MDRL: drug bioavailability; transgenic animal; genetic model.
 XX
 OS Canis familiaris.
 XX
 OS
 FT Key location/Qualifiers
 FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 DR MPI: 2001-235373/24.
 DR N-PSDB; AAD03504.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors.
 XX
 PS Claim 17; Page 91-93; 11pp; English.
 XX
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDRL) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
 CC This sequence is also referred as Genotype A protein. The PGP
 CC enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 CC
 SO Sequence 1281 AA:
 Query Match 12.5%; Score 69; DB 22; Length 1281;
 Best Local Similarity 25.5%; Pred. No. 38;
 Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;
 QY 6 LFFSVITVMMILIMASEMV-----NGSAFTVWSGPCNNNAERYSKCGCSAITHQGYD 59
 DB 336 VFTSVLLGAFSGAASLEAFANARGAAEYELIKLIDNKPISYSKSHKPDNIKGNLE 395
 QY 60 -----FSY-----TGQTAALYNQAC 75
 DB 396 fknvhfsypsrkevklkglnlkvsgqvalvngsgc 433
 RESULT 10
 AAE00309
 ID AAE00309 standard; Protein: 1281 AA.
 XX
 AC AAE00309;

XX	
DT	13-JUN-2001 (first entry)
XX	
DE	Dog P-glycoprotein (PGP) allelic variant (Genotype B).
XX	
KW	Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
XX	MDR1; drug bioavailability; transgenic animal; genetic model.
OS	Canis familiaris.
XX	
FH	Key
FT	Misc-difference 25 Location/Qualifiers
FT	/note= "asn of GenotypeC substituted by Lys"
FT	197
FT	Misc-difference /note= "His of genotypec substituted by Gln"
XX	
PN	WO200123540-A2.
XX	
PD	05-APR-2001.
PF	28-SEP-2000; 2000MO-US26767.
XX	
PR	28-SEP-1999; 99US-0156510.
XX	
PA	(GENT-) GENTEST CORP.
XX	
PI	Stocker RJ, Steinel-crespi DT, Crespi CL, Relif TC, Patten CJ;
XX	
DR	WPI: 2001-235373/24.
XX	N-PSTB; MAD03505.
PT	New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT	for determining the bioavailability of drugs and for screening for dog
XX	PGP inhibitors -
PS	Claim 17; Page 99-102; 11pp; English.
XX	
CC	The invention relates to dog P-glycoprotein (PGP) also referred
CC	as multidrug transporter (MDR1) and nucleic acids encoding them.
CC	The invention also includes fragments and biologically functional
CC	variants of dog P-glycoprotein. PGP and their nucleic acids are
CC	useful for determining the bioavailability of drugs and for
CC	screening PGP inhibitors. They are useful for the diagnosis and
CC	treatment of conditions characterized by PGP activity, by
CC	reducing or increasing PGP activity in a cell. PGP nucleic acids
CC	are used as oligonucleotide probes. Complements of PGP nucleic
CC	acids are useful as antisense oligonucleotides, to induce a PGP
CC	"knockout" phenotype. They are used to prepare a non-human
CC	transgenic animal, which are valuable as genetic models for
CC	human diseases.
CC	The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC	This sequence is also referred as Genotype B protein. The PGP
CC	enzyme functions as an efflux pump exporting small molecules
CC	across the cell membrane. This enzyme is a member of the ABC
CC	transporter family.
XX	
Sequence	1281 AA;
XX	

```
Query Match      12.5% Score 69; DB 22; Length 1281;
Best Local Similarity 25.5%; Pred. No. 38;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3
```

OY 6 LFFSVITWMLLMAASEMW-----NGSAFTWSGEGCCNNRERERSKCCCSAIHOKGYD 59
:||||: | ||| : * |: : : ||| :
DB 336 vtfsvllgafsgdgspsleatfanaargaaeyelfkllldkpsidsysksghkpndnlkgule 395
OY 60 -----FSY-----TCGTALLYNQAGC 75
||| :||| || :||
DB 396 fknvntfsypsrkevklkglmlkvsgsqgtvalvngnsc 433

RESULT 11

ID	AAE00310	
AC	AAE00310 standard; Protein; 1281 AA.	
XX		
XX	AAE00310;	
DT	13-JUN-2001 (first entry)	
XX		
DE	Dog P-glycoprotein (PGP) allelic variant (Genotype D).	
KW	Dog; P-glycoprotein allelic variant; PGP; multidrug transporter; MDRI; drug bioavailability; transgenic animal; genetic model.	
XX		
OS	Canis familiaris.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 25	/note= "Asn of genotpec substituted by Lys"
FT	Misc-difference 197	/note= "His of genotpec substituted by Gln"
FT	Misc-difference 329	/note= "Ser of genotpec substituted by Thr"
FT	Misc-difference 1148	/note= "Met of genotpec substituted by Val"
XX		
PN	WO200123540-A2.	
PD	05-APR-2001.	
XX		
PF	28-SEP-2000; 2000WC-US26767.	
PR	28-SEP-1999; 99US-0156510.	
XX		
PA	(GENT-) GENTEST CORP.	
PI	Stockert PJ, Steimel-Crespi DT, Crespi CL, Relif TC, Patten CJ; WPI: 2001-235373/24.	
DR	N-PSDB; AAD03506.	
PT	New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful for determining the bioavailability of drugs and for screening for dog PGP inhibitors -	
XX		
PS	Claim 17: Page 108-110; 11pp; English.	
XX		
CC	The invention relates to dog P-glycoprotein (PGP) also referred as multidrug transporter (MDRI) and nucleic acids encoding them. The invention also includes fragments and biologically functional variants of dog P-glycoprotein. PGP and their nucleic acids are useful for determining the bioavailability of drugs and for screening PGP inhibitors. They are useful for the diagnosis and treatment of conditions characterised by PGP activity, by reducing or increasing PGP activity in a cell. PGP nucleic acids are used as oligonucleotide probes. Complements of PGP nucleic acids are useful as antisense oligonucleotides, to induce a PGP 'knockout' phenotype. They are used to prepare a non-human transgenic animal, which are valuable as genetic models for human diseases.	
CC	This present sequence is dog P-glycoprotein (PGP) allelic variant.	
CC	This sequence is also referred as Genotype D protein. The PGP enzyme functions as an efflux pump exporting small molecules across the cell membrane. This enzyme is a member of the ABC transporter family.	
XX		
Sequence	1281 AA:	
1Q		

```
Query Match      12.5%; Score 69; DB 22; Length 1281;
Best Local Similarity 25.5%; Pred. No. 38;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

QY    6 LFSVLTVMKLLMAMSEMY-----NSAFTVWSGPGCCNNRREYSKCGCSAIHQKGYD 59
       :|::||: | ||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: |
```

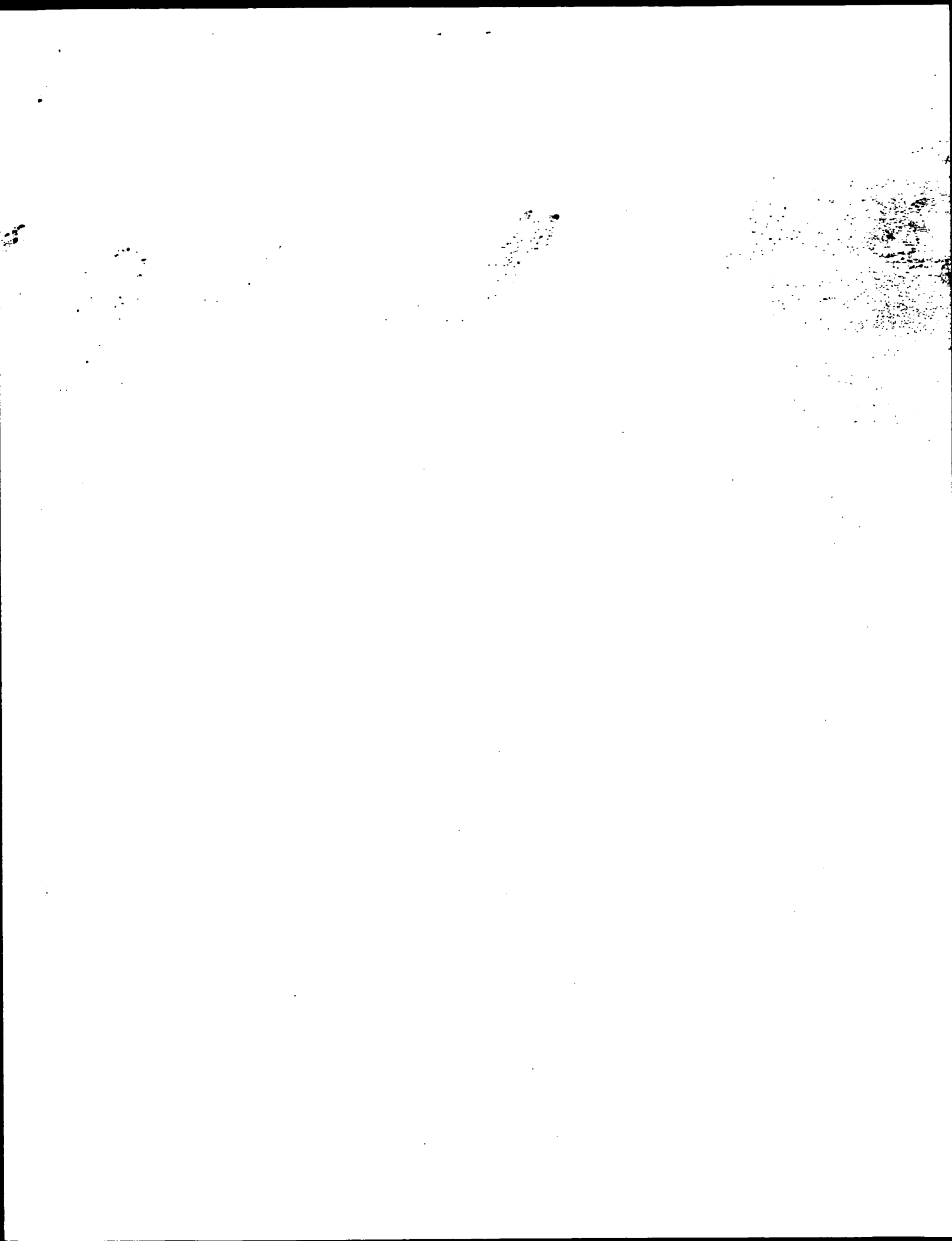

Thu Jan 24 09:57:59 2002

us-09-882-434a-1.rag

Page 9

Db 394 frnvhsypsrkevrllkgjnlkwsqgltvalvgn\$gc 431

Search completed: January 24, 2002, 09:28:22
Job time: 212 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2002, 09:26:40 ; Search time 12.65 seconds
(without alignments)
181.449 Million cell updates/sec

Title: US-09-882-434A-1
Perfect score: 551
Sequence: 1 MASTKLFPSVITVMMILMA.....FGSSARACNPFQKMSIFIQ 102

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing filed: 45 summaries

Database: Issued Patents, AA.*
1: /cgn2-6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2-6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2-6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2-6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2-6/ptodata/2/1aa/PCRTUS.COMB.pep.*
6: /cgn2-6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.5	13.5	486	1 US-07-872-678A-48	Sequence 48, Appl
2	68	12.3	1280	2 US-08-583-276-19	Sequence 19, Appl
3	68	12.3	1280	2 US-08-752-447-2	Sequence 2, Appl
4	68	12.3	1280	6 5206352-4	Patent No. 5206352
5	67.5	12.3	902	1 US-08-701-846-2	Sequence 2, Appl
6	63.5	11.5	1019	1 US-08-296-014A-4	Sequence 4, Appl
7	63.5	11.5	1019	2 US-08-596-405-4	Sequence 4, Appl
8	63.5	11.5	1019	2 US-08-877-620-4	Sequence 4, Appl
9	63.5	11.5	1083	1 US-08-296-014A-2	Sequence 2, Appl
10	63.5	11.5	1083	2 US-08-596-405-2	Sequence 2, Appl
11	63.5	11.5	1083	2 US-08-877-620-2	Sequence 9, Appl
12	62.5	11.3	135	4 US-07-857-224B-99	Sequence 11, Appl
13	62	11.3	3729	2 US-08-804-227C-4	Sequence 4, Appl
14	62	11.3	111	2 US-07-857-224B-105	Sequence 105, App
15	61.5	11.2	135	2 US-07-857-224B-97	Sequence 97, Appl
16	61.5	11.2	135	2 US-07-857-224B-98	Sequence 98, Appl
17	61.5	11.2	1186	1 US-08-144-121-4	Sequence 4, Appl
18	61	11.1	1186	1 US-08-735-893-4	Sequence 4, Appl
19	61	11.1	1186	1 US-08-784-649A-2	Sequence 2, Appl
20	61	11.1	1279	2 US-08-680-326-41	Sequence 41, Appl
21	60.5	11.0	985	2 US-08-612-734B-2	Sequence 2, Appl
22	60.5	11.0	1349	2 US-08-470-702-10	Sequence 10, Appl
23	60	10.9	687	1 US-08-467-831-10	Sequence 10, Appl
24	60	10.9	687	1 US-08-696-944-2	Sequence 2, Appl
25	60	10.9	6730	2 US-08-245-041-19	Sequence 19, Appl
26	60	10.9	671	4 US-09-245-041-19	Sequence 19, Appl
27	60	10.9	1498	4 US-09-245-041-131	Sequence 131, App

28	60	10.9	1350	4 US-09-245-041-17	Sequence 17, Appl
29	60	10.9	2787	4 US-09-245-041-15	Sequence 15, Appl
30	59.5	10.8	305	4 US-09-230-222-1	Sequence 1, Appl
31	59.5	10.8	861	4 US-08-960-046-12	Sequence 12, Appl
32	59	10.7	352	2 US-08-483-926A-11	Sequence 11, Appl
33	59	10.7	549	4 US-09-245-041-9	Sequence 9, Appl
34	59	10.7	687	1 US-08-204-656B-10	Sequence 10, Appl
35	59	10.7	1260	4 US-09-245-041-2	Sequence 2, Appl
36	59	10.7	1345	2 US-08-977-767-3	Sequence 3, Appl
37	58.5	10.6	1404	2 US-08-400-159-2	Sequence 2, Appl
38	58.5	10.6	1404	3 US-08-611-728A-2	Sequence 2, Appl
39	58.5	10.6	2594	4 US-08-718-388-7	Sequence 7, Appl
40	58.5	10.6	5405	4 US-08-718-388-9	Sequence 9, Appl
41	58	10.5	600	3 US-09-212-971-12	Sequence 12, Appl
42	58	10.5	600	4 US-08-800-929A-12	Sequence 12, Appl
43	58	10.5	600	4 US-09-617-053A-12	Sequence 12, Appl
44	57.5	10.4	305	1 US-08-090-013-2	Sequence 2, Appl
45	57.5	10.4	305	1 US-08-081-328-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-07-872-678A-48
; Sequence 48, Application US/07872678A
; Patent No. 5541060
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme, et al.
; TITLE OF INVENTION: DETECTION OF EARLY-ONSET
; TITLE OF INVENTION: NON-INSULIN-DEPENDENT DIABETES MELLITUS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: Post Office Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,678A
; FILING DATE: 22-APRIL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: ARCD016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; US-07-872-678A-48
Query Match 13.5%; Score 74.5; DB 1; Length 486;
Best Local Similarity 39.3%; Pred. No. 0.9;
Matches 24; Conservative 4; Mismatches 18; Indels 15; Gaps 4;
QY 40 RAERYSKGCSAIIHOKGIDSYTGQTRALYNQAGCGVAHTPR-GSSARACNP---FG 94
DB 391 RAARLSVCGIAIACOKRGYK---TGHIAA-----DGSVSTRPGKREKAAALADITG 440

OY 95 W 95
Db 441 W 441

RESULT 2

US-08-583-276-19
Sequence 19, Application US/08583276
Patent No. 5837536
GENERAL INFORMATION:
APPLICANT: McDonagh, Kevin T.
APPLICANT: Nienhuis, Arthur
APPLICANT: Tolstoshev, Paul
TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gillfillan,
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: DM4 V2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/583,276
FILING DATE: 05-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/332,444
FILING DATE: 31-OCT-1994
APPLICATION NUMBER: 07/887,712
FILING DATE: 22-MAY-1992
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: protein
US-08-583-276-19

Query Match 12.3%; Score 68; DB 2; Length 1280;
Best Local Similarity 24.5%; Pred. No. 17;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVITVWMLIMASENV-----NGSAFTVWSPGCGNNRAERYSKCGCSAIHOKGYD 59
Db 334 VFFSVLIGAFSGVQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSGKHPDNINGNLE 393
OY 60 -----FSY-----TGOTAAALVNOAGC 75
Db 394 FRNVHFSYPSRKEVKILKGLMLKVQSGQVVALVGNNGC 431

RESULT 3

US-08-752-447-2
Sequence 2, Application US/0852447
Patent No. 5994088
GENERAL INFORMATION:
APPLICANT: Mechtner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and
TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
STREET: 300 South Wacker Drive, Seventh Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,447
FILING DATE: 15-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5994088nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-9808
TELEFAX: 312-913-0001
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-447-2

Query Match 12.3%; Score 68; DB 2; Length 1280;
Best Local Similarity 24.5%; Pred. No. 17;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVITVWMLIMASENV-----NGSAFTVWSPGCGNNRAERYSKCGCSAIHOKGYD 59
Db 334 VFFSVLIGAFSGVQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSGKHPDNINGNLE 393
OY 60 -----FSY-----TGOTAAALVNOAGC 75
Db 394 FRNVHFSYPSRKEVKILKGLMLKVQSGQVVALVGNNGC 431

RESULT 4

US-08-5206352-4
Patent No. 5206352
APPLICANT: Roninson, Igor B.; Pastan, Ira H.; Gottesman,
Michael M.
TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/622,836
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 892,575
FILING DATE: 01-AUG-1986
APPLICATION NUMBER: 845,610
FILING DATE: 28-MAR-1986
SEQ ID NO: 4
LENGTH: 1280
5206352-4

Query Match 12.3%; Score 68; DB 6; Length 1280;
Best Local Similarity 24.5%; Pred. No. 17;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;
OY 6 LFFSVITVWMLIMASENV-----NGSAFTVWSPGCGNNRAERYSKCGCSAIHOKGYD 59
Db 334 VFFSVLIGAFSGVQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSGKHPDNINGNLE 393
OY 60 -----FSY-----TGOTAAALVNOAGC 75
Db 394 FRNVHFSYPSRKEVKILKGLMLKVQSGQVVALVGNNGC 431

REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-596-405-4

Query Match 11.5%; Score 63.5; DB 2; Length 1019;
Best Local Similarity 32.8%; Pred. No. 43;
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

QY 32 WSG--PCNNRAERYSCGSAIHQKGYDF-SYTGOTALYNQACSGVAHTRFGSS 86
DB 314 WSGSPSCVAVADREVDCSKAV-----DFLDVGEVPVRIHCPAGCSLTAGTVMGTA 365

RESULT 8

US-08-877-620-4
Sequence 4, Application US/0877620
Patent No. 5985590
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
APPLICANT: Ho, Bow
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carinoscorpius
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877,620
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/596,405
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-877-620-4

Query Match 11.5%; Score 63.5; DB 2; Length 1019;
Best Local Similarity 32.8%; Pred. No. 43;

Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;
QY 32 WSG--PCNNRAERYSCGSAIHQKGYDF-SYTGOTALYNQACSGVAHTRFGSS 86
DB 314 WSGSPSCVAVADREVDCSKAV-----DFLDVGEVPVRIHCPAGCSLTAGTVMGTA 365

RESULT 9

US-08-296-014A-2
Sequence 2, Application US/08296014A
Patent No. 5716834
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
APPLICANT: Ho, Bow
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carinoscorpius
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,014A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1083 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-296-014A-2

Query Match 11.5%; Score 63.5; DB 1; Length 1083;
Best Local Similarity 32.8%; Pred. No. 46;
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

QY 32 WSG--PCNNRAERYSCGSAIHQKGYDF-SYTGOTALYNQACSGVAHTRFGSS 86
DB 378 WSGSPSCVAVADREVDCSKAV-----DFLDVGEVPVRIHCPAGCSLTAGTVMGTA 429

RESULT 10

US-08-596-405-2
Sequence 2, Application US/08596405
Patent No. 5858706
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
APPLICANT: Ho, Bow
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carinoscorpius
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: 8110 Gatehouse Road, Suite 500 East
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22042
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/596,405
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1781-105P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 205-8000
 TELEFAX: (703) 205-8050
 TELEX: 248345
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1083 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-596-405-2

Query Match 11.5%; Score 63.5; DB 2; Length 1083;
 Best Local Similarity 32.8%; Pred. No. 46;
 Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;
 QY 32 MSG--PGCNNAERYSKGCSAIHQKGYDF-SYTGQTALYNQAGCGVAHTREGSS 86
 DB 378 MSGSQPCVKVADREVDCSKAV-----DFLDVGEPRVHCPCAGCSLTAGTVWGTA 429

RESULT 11
 US-08-877-620-2
 Sequence 2, Application US/08877620
 Patent No. 5985590
 GENERAL INFORMATION:
 APPLICANT: Ding, Jeak Ling
 APPLICANT: Ho, Bow
 TITLE OF INVENTION: The Cloned Factor C cDNA of the
 TITLE OF INVENTION: Singapore Horseshoe Crab, Carlinoscorpius
 TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch & Birch
 STREET: 8110 Gatehouse Road, Suite 500 East
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22042
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/877,620
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/596,405
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Jr., Gerald M.
 REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 205-8000
 TELEFAX: (703) 205-8050
 TELEX: 248345
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1083 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-877-620-2

Query Match 11.5%; Score 63.5; DB 2; Length 1083;
 Best Local Similarity 32.8%; Pred. No. 46;
 Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;
 QY 32 MSG--PGCNNAERYSKGCSAIHQKGYDF-SYTGQTALYNQAGCGVAHTREGSS 86
 DB 378 MSGSQPCVKVADREVDCSKAV-----DFLDVGEPRVHCPCAGCSLTAGTVWGTA 429

RESULT 12
 US-07-857-224B-99
 Sequence 99, Application US/07857224B
 Patent No. 5958784
 GENERAL INFORMATION:
 APPLICANT: Benner, Steven A.
 TITLE OF INVENTION: Predicting Folded Structures of Proteins
 NUMBER OF SEQUENCES: 114
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Steven A. Benner
 STREET: Hadlaubstrasse 151
 CITY: Zurich
 STATE: none
 COUNTRY: Switzerland
 ZIP: (note: this is an international post code) CH-8092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
 OPERATING SYSTEM: Macintosh
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/857,224B
 FILING DATE: 03/25/92
 CLASSIFICATION: 436
 PRIOR APPLICATION DATA: none
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (international) 41 1 632 2830
 TELEFAX: (international) 41 1 262 2437
 TELEX: none
 INFORMATION FOR SEQ ID NO: 99:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 135
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE:
 DESCRIPTION: protein
 ORIGINAL SOURCE:
 ORGANISM: Lycopersicon esculentum
 FEATURE: Pathogenesis related protein; Table 16 Row 3
 PUBLICATION INFORMATION:
 AUTHORS:
 AUTHORS: Cutt, J. R.
 AUTHORS: Dixon, D. C.
 AUTHORS: Carr, J. P.
 AUTHORS: Klessig, D. F.
 TITLE: Isolation and nucleotide sequence of cDNA clones for the
 TITLE: pathogenesis related proteins of Nicotiana tabacum induced by TMV
 TITLE: infection.
 JOURNAL: Nucleic Acids Research
 VOLUME: 16

PAGES: 9861
DATE: 1988
US-07-857-224B-99

Query Match 11.3% Score 62.5; DB 2; Length 135;
Best Local Similarity 31.2% Pred. No. 4.6;
Matches 30; Conservative 5; Mismatches 24; Indels 37; Gaps 9;

OY 39 NRAERY--SKG-CSAIHQ-----KGYDESYTGOTAL-----YNOA--GCSG- 77
DB 31 SRAQVYANSRAGDCMLHSGGEMLANGGDF--TGRAVQVWVSERPDYVATNQCVG 88
OY 78 --VAH-----TRGSSARACNPGKXSIPTQC 102
DB 89 KMGCHYTOVWVRNSVRLCGGRACNNGW---FLSC 121

RESULT 13
US-09-347-833-11
Sequence 11, Application US/09347833
Patent No. 6294658
GENERAL INFORMATION:
APPLICANT: Farnodu, Lajo O.
TITLE OF INVENTION: Factors Involved in Gene Expression
FILE REFERENCE: BB-1172
CURRENT APPLICATION NUMBER: US/09/347,833
EARLIER FILING DATE: 1999-07-02
CURRENT APPLICATION NUMBER: 60/092,415
EARLIER FILING DATE: July 10, 1998
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Microsoft Office 97
SEQ ID NO 11
LENGTH: 442
TYPE: PRT
ORGANISM: Trilicium aestivum
US-09-347-833-11

Query Match 11.3% Score 62; DB 4; Length 442;
Best Local Similarity 25.9% Pred. No. 23;
Matches 21; Conservative 9; Mismatches 43; Indels 8; Gaps 2;
OY 14 MMLIAMSSEVNGSAFTWMSGPGCNRAERYSKGCSAIHQKGYDESYTGOTALYNOA 73
DB 326 MMWYGRAGRAGAGMSMGMGGRGGRGYRTGRGMDAMHGVGVTSTYMG---GMMRDA 382
OY 74 GCGSVATHRRGSSARACNPG 94
DB 383 GSVGA-----SANSTRMKG 398

RESULT 14
US-08-804-227C-4
Sequence 4, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostack, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. HANF, 1501
STREET, LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3729 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-4

Query Match 11.3% Score 62; DB 2; Length 3729;
Best Local Similarity 28.8% Pred. No. 3.2e+02;
Matches 21; Conservative 10; Mismatches 34; Indels 8; Gaps 3;
OY 24 VNSAFTVMSG-PGCGNNRERSKCGCSAIHQKGYDESYTGOTALYNOACSGVAH-- 80
DB 1343 VYSVAWTPMLGTPPAADSLQMSRAGITPLDPAASLD---ALARA VGRACCVYADID 1398
OY 81 -TRGSSARACNP 92
DB 1399 WERFASATYATRP 1411

RESULT 15
US-07-857-224B-105
Sequence 105, Application US/07857224B
Patent No. 5958784
GENERAL INFORMATION:
APPLICANT: Benner, Steven A.
TITLE OF INVENTION: Predicting Folded Structures of Proteins
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: Hadlaubstrasse 151
CITY: Zurich
STATE: none
COUNTRY: Switzerland
ZIP: (note: this is an international post code) CH-8092
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/857,224B
FILING DATE: 03/25/92
CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (international) 41 1 632 2830
TELEFAX: (international) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 111
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: protein
ORIGINAL SOURCE: tobacco
ORGANISM: tobacco

```

FEATURE: Pathogenesis related protein; Table 17 Row 1
PUBLICATION INFORMATION:
AUTHORS:
AUTHORS: Curt, J. R.
AUTHORS: Dixon, D. C.
AUTHORS: Carr, J. P.
AUTHORS: Klessig, D. F.
TITLE: Isolation and nucleotide sequence of cDNA clones for the
TITLE: pathogenesis related proteins of Nicotiana glauca induced by TMV
TITLE: Infection.
JOURNAL: Nucleic Acids Research
VOLUME: 16
PAGES: 9861
DATE: 1988
US-07-857-224B-105

```

```

Query Match      11.28; Score 61.5; DB 2; Length 111;
Best Local Similarity 30.28; Pred. No. 4.7;
Matches 29; Conservative 5; Mismatches 25; Indels 37; Gaps 8;

QY 39 NRAERY--SKCG-CSATHQ-----KGYDFSYTGQTAA-----LYNOA-----GC 75
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9 SRQNTANSRAGDCLHSGAGENLAKGGDF--TGRAVQLWVSERPSYVATNOCVGG 66

QY 76 SGVAH-----TRFGSSARACNPFQWKSIFIOC 102
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 KCRHYTQVWVRNSVRLGCGRCARCNNGW---FLSC 99

```

Search completed: January 24, 2002, 09:28:42
 Job time: 122 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2002, 09:27:00 ; Search time 14.39 Seconds
(without alignments)
539.945 Million cell updates/sec

Title: US-09-882-434A-1

Perfect score: 551

Sequence: 1 MASTKLFESVITVMMILAMA.....FGSSARACNPFGRKSIPIQC 102

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries:

Database :
1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.5	13.7	315	2	T35804
2	75.5	13.7	486	1	KIRYHB
3	69	12.5	1281	2	I48123
4	68.5	12.4	1557	2	T28811
5	68	12.3	1280	1	DVHU1
6	67.5	12.3	184	2	T16044
7	66.5	12.1	151	2	T02667
8	66.5	12.1	151	2	JG6040
9	66.5	12.1	157	2	S31078
10	66.5	12.1	162	2	S21157
11	66.5	12.1	251	2	T34656
12	66	12.0	1276	1	DVMS2
13	65.5	11.9	233	2	T22396
14	65.5	11.9	244	2	T04212
15	65	11.8	165	2	S31080
16	64.5	11.7	485	1	KIRYHA
17	64	11.6	320	2	S00935
18	64	11.6	1278	2	S41646
19	64	11.6	1609	1	MMHUB2
20	64	11.6	3016	2	S77300
21	63.5	11.5	166	2	S31082
22	63.5	11.5	503	2	S31940
23	63.5	11.5	828	1	D35142
24	63.5	11.5	1019	2	A38738
25	63	11.4	157	2	T02664
26	63	11.4	159	2	S26238
27	63	11.4	339	2	A47611
28	63	11.4	1276	1	DVHR1C
29	63	11.4	1279	1	DVHU3

30	62.5	11.3	125	2	T16247	hypothetical prote
31	62.5	11.3	166	2	S59922	allergen RA14B pre
32	62.5	11.3	168	2	T07146	pathogenesis-relat
33	62.5	11.3	235	1	QTT2	linactin II precu
34	62.5	11.3	779	2	H71301	pathogenesis-relat
35	62	11.3	159	1	VCT014	heat shock 18k pro
36	62	11.3	169	2	S00900	hypothetical prote
37	62	11.3	234	2	T21495	NADH dehydrogenase
38	62	11.3	340	2	T13781	hypothetical prote
39	62	11.3	553	2	T52362	probable membrane
40	61.5	11.2	455	2	S46033	surface protein ty
41	61.5	11.2	2395	1	S50820	hydroxymethylgluta
42	61	11.1	180	2	B83064	ammonium transport
43	61	11.1	501	2	S59946	hydroxymethylgluta
44	61	11.1	501	2	T06653	
45	61	11.1	574	2	T07112	

ALIGNMENTS

RESULT 1
T35804
Probable aldehyde dehydrogenase - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35804
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Randal, M.A.
Submitted to the EMBL Data Library, August 1998
A:Reference number: Z21556
A:Accession: T35804
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-315 <MUR>
A:Cross-references: EMBL:AL031225; PIDN:CAA20223.1; GSPDB:GN00070; SCOEDB:SC8B7.12C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8B7.12C

Query Match 13.7%; Score 75.5; DB 2; Length 315;
Best Local Similarity 25.9%; Pred. No. 1.8;
Matches 30; Conservative 20; Mismatches 47; Indels 19; Gaps 5;

QY 1 MASTKLFESVITVMMILAMA-----IMASEMNGSAFTVSGRCN-NRAERSKGGCSAII 53
DB 199 VQSEIFEPVLVPEFTDDEGIRLANDTPYGLASAWSRDYRANRATREIKAGCWIN 258

QY 54 Q-----KGYDFSYTGOTAAALYNQAGCGVAHFRFGSSARACNPFGR-KSIF 99
DB 259 DIIPISEMPHGGYKASGFGKMSAYSFEEYTKVKNVEMDTAVAAK--DMHRTVF 312

RESULT 2
KIRYHB
hexokinase (EC 2.7.1.1) B - yeast (Saccharomyces cerevisiae)
A:Alternate names: HEX1 protein; hexokinase II; hexokinase PII; protein G0556; protei
C:Species: Saccharomyces cerevisiae
C:Date: 28-Dec-1987 #sequence_revision 12-Apr-1996 #text_change 16-Jun-2000
C:Accession: S61608; B23523; S22430; A23958; S05731; S33656; S64279; A53632; S28555
R:Coissac, E.; Maillet, E.; Robineau, S.; Neller, P.
Submitted to the EMBL Data Library, December 1995
A:Reference number: S61598
A:Accession: S61608
A:Molecule type: DNA
A:Residues: 1-486 <COI>
A:Cross-references: EMBL:X94357; MID:g1150575; PIDN:CAA64134.1; PID:g1150586
R:Stachek, C.; Stachek, J.; Swan, J.; Botstein, D.; Konigsberg, W.
Nucleic Acids Res. 14, 945-963, 1986
A:Title: Identification, cloning and sequence determination of the genes specifying h
A:Reference number: A93649; M01D:66120382
A:Accession: B23523
A:Molecule type: DNA

A:Residues: 1-28,'I',30-60,'V',62-196,'S',198-486 <STAS>
A:Cross-references: EMBL:X03483; NID:q3792; PIDD:CAA27203.1; PID:q3793
A:Note: the authors translated the codon GTT for residue 61 as Gly
A:Accession: S22430
A:Molecule type: protein
A:Residues: 2-7,'I',30-46;78-111;114-119;166-173;177-194;228-236;250-281;283-300;330-335
R:Frohlich, K.; Elik, K.; Mecke, D.
Gene 36, 105-111, 1985
A:Title: The primary structure of the yeast hexokinase PII gene (HXK2) which is responsible
A:Reference number: A23958; MUID:8605643
A:Accession: A23958
A:Molecule type: DNA
A:Residues: 1-32,'N',34-60,'V',62-420,'ST',423-443,'PH',446-452,'V',454-461,'P',463-486
A:Cross-references: EMBL:M1181
A:Note: the authors translated the codon GTT for residue 61 as Gly
R:Schmidt, J.J.; Colowick, S.P.
Arch. Biochem. Biophys. 158, 471-477, 1973
A:Title: Identification of a peptide sequence involved in association of subunits of yeast
A:Reference number: S05731; MUID:74114889
A:Accession: S05731
A:Molecule type: protein
A:Residues: 2-12 <SCH>
R:Breitwieser, W.; Price, C.; Schuster, T.
Yeast 9, 551-556, 1993
A:Title: Identification of a gene encoding a novel zinc finger protein in Saccharomyces
A:Reference number: S33654; MUID:9311123
A:Accession: S33656
A:Molecule type: DNA
A:Residues: 1-247 <BRE>
A:Cross-references: EMBL:X67787; NID:q3707; PIDD:CAA48003.1; PID:g3710
R:Collaço, E.; Maillat, E.; Netter, P.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64271

```

Query Match 13.7%; Score 75.5; DB 1; Length 486;
Best Local Similarity 30.2%; Pred. No. 2.6;
Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;

Oy 8 FSVITVMM---LITAMSEWNGSAFTWVSGPCNNNAERYSKCGCAIHQKGYDESYNG 64
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 369 FGINPTVQERKLIRLSELTIGA-----NAAFLSVCGTALICQKRGYK---TG 412
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 65 QTAALYNQAGCGVAHTRE-GSSARACNP---FGM 95
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 413 HTAA-----DGSYINRYGCEKREKAAALKDITGM 441
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
148123
P-glycoprotein isoform IML-1-Chinese hamster

```

C:Species: *Cricketulus griseus* (Chinese hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
C:Accession: I48123
R:Endicott, J. A.; Sarangl, F.; Ling, V.
DNA Seq. 2, 89-101, 1991
A:Title: Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene fam
A:Reference numbers: I48121; M01D:92135896
A:Accession: I48123
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1281 <RCS>
A:Cross-references: GB:M0042; NID:g191168; PIDN:AAA68885.1; PID:g191169
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop
F:413-606/Domain: ATP-binding cassette homology <AC1>
F:432-436/Region: nucleotide-binding motif A (P-loop)
F:1054-1250/Domain: ATP-binding cassette homology <ABC>
F:1071-1078/Region: nucleotide-binding motif A (P-loop)

```

Query Match      12.5%; Score 69; DB 2; Length 1281;
Best Local Similarity 21.4%; Pred. No. 31;
Matches 21; Conservative 17; Mismatches 32; Indels 28; Gaps 2

OY      6  LFEVITYVMHLMAMSMVN-----GSAFYWSGPCGNNRERYSGCCGAIHQKGYD 59
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      336  VEFSTILIGASVGQAPCIDAFANRAGARVLEFDIIDNNPKIDSFSENGKRPDSIKGNLD 395
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      60  FS-----YTGQTALYNOAGC 75
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      396  FSDVHFSYSPRANIKILKGLINLKVGQGTFFALVANGSGC 433
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 4
T28811
hypothetical protein C54D1.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence:revliston 29-Oct-1999 #ext_change 18-Feb-2000
C:Accession: J28811
R:Minx, M.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid C54D1.
A:Reference number: 220527
A:Accession: T28811
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1557 <MIN>
A:Cross-references: EMBL:U046673; PIDN:MAC48152.1; GSPDB:GN00028; CESP:C54D1.5
A:Experimental source: strain Bristol N2; clone C54D1
C:Genetics:
A:Gene: CESP:C54D1.5
A:Map position: X
A:Introns: 84/3; 127/1; 166/2; 271/2; 331/1; 392/3; 433/2; 585/2; 1089/1; 1530/3
C:Superfamily: lamtln beta-1 chain; lamtln-type EGF-like homology

```

```

Query Match Similarity      12.4%  Score 68.5:  Db 2:  Length 1557;
Best Local Similarity      28.8%  Pred No. 41;
Matches 19;  Conservative  5;  Mismatches 33;  Indels  9;  Gaps  2;

QY      29  FTVWSGPGCCNNBAERYSKGCSAIIHQKGGYDFSYTGQTALYNOAG--CSGYAHITRFSSS  86
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       920  YNITSGLG-----QECNCDPPLASEGNTCDVNTGCCCKPGVYIGORCDRCADYHFGFS  972

QY      87  ARACNP  92
      |  |  |
Db       973  ANGCP  978

RESULT      5
DYNHUI
multidrug resistance protein 1 - human
N:Alternate names: P-glycoprotein 1

```

C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1990 #sequence_revision: 18-Aug-1995 #text_change: 19-Jan-2001
 C:Accession: A34914; PS01627; S15500; S43838; I52238; I65204
 R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.
 J. Biol. Chem. 265: 506-514, 1990
 A:Title: Genomic organization of the human multidrug resistance (MDR1) gene and origin
 A:Reference number: A34914; MUID:90094448
 A:Molecule type: DNA
 A:Accession: A34914
 A:Residues: 1-1280 <CHP>
 A:Cross-references: GB:M29447; GB:J05168; NID:g187496; PIDN:AAA59576.1; PID:g386862
 R:Kloke, N.; Yamano, Y.; Komano, T.; Ueda, K.
 submitted to JIPID, April 1991
 A:Reference number: PS0162
 A:Accession: PS0162
 A:Molecule type: DNA
 A:Residues: 1-22 <KIO>
 R:Kloke, N.; Yamano, Y.; Komano, T.; Ueda, K.
 submitted to the EMBL Data Library, April 1991
 A:Description: Transcriptional regulation of multidrug resistance gene (MDR1) expression
 A:Reference number: S15500
 A:Accession: S15500
 A:Molecule type: DNA
 A:Residues: 1-22, 'R' <KIO>
 A:Cross-references: EMBL:X58723; NID:g34522; PIDN:CAA41558.1; PID:g34523
 R:Chen, C.; Chiu, J.E.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.
 Cell 47, 381-389, 1986
 A:Title: Internal duplication and homology with bacterial transport proteins in the mdr1
 A:Reference number: A25059; MUID:87028230
 A:Accession: A25059
 A:Molecule type: mRNA
 A:Residues: 1-184, 'V', 186-1280 <CH2>
 A:Cross-references: GB:M14758; NID:g187468; PIDN:AAA59575.1; PID:g307180
 R:Chambers, T.C.; Pohl, J.; Gifford, D.B.; Kuo, J.F.
 Biochem. J. 299, 309-315, 1994
 A:Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase of
 A:Reference number: S43838; MUID:94220047
 A:Accession: S43838
 A:Molecule type: protein
 A:Residues: 656-689 <CHN>
 R:Geckeler, V.; Weger, S.; Probst, H.
 Biochem. Biophys. Res. Commun. 169, 796-802, 1990
 A:Title: mdr1/P-glycoprotein gene segments analyzed from various human leukemic cell lit
 A:Reference number: I52238; MUID:90280529
 A:Accession: I52238
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 178-215 <RES>
 A:Cross-references: GB:M37724; NID:g183537; PIDN:AAA88047.1; PID:g553314
 A:Accession: I65204
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 800-856 <RE2>
 A:Cross-references: GB:M37725; NID:g183538; PIDN:AAA88048.1; PID:g553315
 C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
 structurally and functionally unrelated lipophilic antitumor drugs.
 C:Genetics:
 A:Gene: GDB:PG1; MDR1
 A:Cross-references: GDB:120712; OMIM:171050
 A:Map position: 7q21-7q21
 C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
 C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; phosphoprotein;
 F:1-658,653-1280/Region: duplication
 F:49-350/Domain: hydrophobic <HBI>
 F:351-637/Domain: hydrophobic <HBI>
 F:410-604/Domain: ATP-binding cassette homology <ABC1>
 F:427-434/Region: nucleotide-binding motif A (P-loop)
 F:551-555/Region: nucleotide-binding motif B
 F:638-708/Domain: linker <LIN>
 F:709-993/Domain: hydrophobic <HB2>
 F:994-1280/Domain: hydrophobic <HL2>
 F:1053-1249/Domain: ATP-binding cassette homology <ABC2>
 F:1070-1077/Region: nucleotide-binding motif A (P-loop)

F:1196-1200/Region: nucleotide-binding motif B
 F:91,94,99/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:433/Binding site: ATP (lys) #status predicted
 F:661,667,671/Binding site: phosphate (ser) (covalent) (by protein kinase C) #status
 F:667,671,683/Binding site: phosphate (ser) (covalent) (by cAMP-dependent kinase) #st
 F:1076/Binding site: ATP (lys) #status predicted

Query Match 12.3% Score 68; DB 1; Length 1280;
 Best Local Similarity 24.5% Pred. NO. 39;
 Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;
 QY 6 LPSFTVVMILAMSENV-----NSAFTWSPGCGNNRAERYSKCSAIIHQGGYD 59
 DB 334 VFTSVLIGAFSVQASPSLEAFANRGAAYEIFKIDDKPSIDSYSKGHPDMIKMLE 393
 QY 60 ----FSY-----TGOTALLYNOAC 75
 DB 394 FRNVHFSYRSRKEVKILKLNKLVOSGOTVALVNSGC 431

RESULT 6
 T16044
 hypothetical protein F12A10.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision: 20-Sep-1999 #text_change: 20-Sep-1999
 R:Geisler, C.
 submitted to the EMBL Data Library, June 1995
 A:Description: The sequence of C. elegans cosmid F12A10.
 A:Reference number: Z18451
 A:Accession: T16044
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-184 <GEI>
 A:Cross-references: EMBL:U28731; NID:g861241; PID:g861242; PIDN:AAA68294.1; CESP:F12A
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F12A10.1
 A:Introns: 25/2; 66/1; 114/2

Query Match 12.3% Score 67.5; DB 2; Length 184;
 Best Local Similarity 24.1% Pred. No. 7.2; 34; Indels 37; Gaps 4;
 Matches 26; Conservative 11; Mismatches
 QY 1 MASRTLFPSVITVMMILAMSENV-----GSAFTWMSG-----PG 36
 DB 78 MHSTLTFEYVFGILVNSIAFDLNNHKKLNQWSSSEGRVKKRMGGGPGYGGCGYRG 137
 QY 37 CNNAERYSKCSAIIHQGGYDFSYTGOTALLYNOACGCVAHTRPG 84
 DB 138 -----GYSGYG-----GGYPMYGGCMGSGSYSSMSGSSSSRSRG 172

RESULT 7
 T02667
 proteinase inhibitor - rice
 C:Species: Oryza sativa (rice)
 C>Date: 24-Mar-1999 #sequence_revision: 24-Mar-1999 #text_change: 09-Jun-2000
 R:Yun, C.H.; Lee, J.H.; Park, J.H.; Lee, G.R.
 submitted to the EMBL Data Library, January 1998
 A:Description: Molecular characterization of rice proteinase inhibitor gene.
 A:Reference number: Z14694
 A:Accession: T02667
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-81 <YUN>
 A:Cross-references: EMBL:AF044059; NID:g2829211; PIDN:AAC00503.1; PID:g2829212
 A:Experimental source: strain IR36
 C:Genetics:
 A:Gene: RGP19

A:Introns: 21/1
C:Superfamily: gamma-thionin

Query Match

Best Local Similarity 12.1%; Score 66.5; DB 2; Length 81;
Matches 19; Conservative 16; Mismatches 28; Indels 19; Gaps 3;

OY 2 ASKLFESVITVMMLIMASEMNGSAFTWSGPGCNRAERY-----KCGSAIHQK 56
DB 3 ASKVFSAFMLVLLATGEM--GCPVMAEARTECSOSHRKCPKARKANCAVCNTE 60
OY 57 GYDSYTGQTALYNOAGCSGV 78
DB 61 GEPDGY-----CHGV 70

RESULT 8

JC6040

fimbriin protein agfB precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: JC6040

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A:Reference number: JC6039; MUID:96146512

A:Accession: JC6040

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:91184712; PIDN:AAC43598.1; PID:91184713

A:Experimental source: strain 276755-3b

C:Genetics:

A:Gene: agfB

C:Function:

A:Description: minor component of thin aggregative fimbriae

A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: fimbriin protein agfB #status predicted <MAT>

Query Match 12.1%; Score 66.5; DB 2; Length 151;
Best Local Similarity 31.0%; Pred. No. 7.6;
Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;

OY 30 TWSGCGCNRAERYSKCCSAIHQKGYDSYTGQTALYNOAGCSGVATTRGSSA 87
DB 69 SVISGCGNNRAK-----VDQAGNYNFAIYEQT-----GNANDASISQSAVGNSSA 113

RESULT 9

S31078

seed allergen RA5 - rice

C:Species: Oryza sativa (rice)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000

C:Accession: S31078

R:Adachi, T.; Izumi, T.; Yamada, T.; Tanaka, K.; Takeuchi, S.; Nakamura, R.; Matsuda, T.

Plant Mol. Biol. 21, 239-248, 1993

A:Title: Gene structure and expression of rice seed allergenic proteins belonging to the

A:Reference number: S31078; MUID:93144699

A:Accession: S31078

A:Molecule type: mRNA

A:Residues: 1-157 <ADA>

A:Cross-references: EMBL:D11430; NID:9218196; PIDN:BA01996.1; PID:9218197

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: seed

Query Match 12.1%; Score 66.5; DB 2; Length 157;
Best Local Similarity 22.1%; Pred. No. 7.9;
Matches 33; Conservative 12; Mismatches 35; Indels 69; Gaps 7;

OY 1 MASTKLFESV--ITVMMLIMASEMNV-----GSAFTWSGPGCNRAE 42
DB 1 MASNKVFESVLLAVVSVLAATATMAEYHHQDVVTBARCGPGMGVPMYSLPRCALVK 60
OY 43 RY-----SKCCSAI-HQKGYDSYTGQTALYNOAGCSG 77
DB 61 RCGGSAAAAEQVRRDCCRLAAVDDSCRCCEAISHMLG-----TYRELGAVD 109
OY 78 VAH-----TRFGSARA-----CN 91
DB 110 VGHPMSEVFERGCRGDLERAASLPAFCN 138

RESULT 10

S21157

seed allergen RA17 - rice

C:Species: Oryza sativa (rice)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000

C:Accession: S21157; JC4887; S31079

R:Izumi, H.; Adachi, T.; Fujii, N.; Matsuda, T.; Nakamura, R.; Urisu, A.;

FEBS Lett. 302, 213-216, 1992

A:Title: Nucleotide sequence of a cDNA clone encoding a major allergenic protein in r

A:Reference number: S21157; MUID:92289999

A:Accession: S21157

A:Molecule type: mRNA

A:Residues: 1-162 <IZU>

A:Cross-references: EMBL:D11431; NID:9218194; PIDN:BA01997.1; PID:9218195

R:Nakamura, R.; Matsuda, T.

Biosci. Biotechnol. Biochem. 60, 1215-1221, 1996

A:Title: Rice allergenic protein and molecular-genetic approach for hypoallergenic r

A:Reference number: JC4887; MUID:97141195

A:Contents: seed

A:Accession: JCA887

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-162 <NAK>

A:Cross-references: EMBL:X66257; NID:9311892; PIDN:CAA46983.1; PID:9311893

C:Comment: This protein shows a sequence similarity to alpha-amylase/trypan inhibitor

Query Match 12.1%; Score 66.5; DB 2; Length 162;
Best Local Similarity 21.5%; Pred. No. 8.2;
Matches 31; Conservative 17; Mismatches 37; Indels 59; Gaps 7;

OY 1 MASTKLFESVITVMMLIMASEMNV-----GSAFTWSGPGCNRAERY 44
DB 1 MASNKVFESVLLAVVSVLAATATMAEYHHQDVVTBARCGPGMGVPMYSLPRCALVK 59
OY 45 SKC-----GCSAIHQK-----GGYDSYTGQTALYNOAGCSGVAA-- 80
DB 60 -OCVGRGASAADEQVQDCRQLAAVDDGWCRCGALDHMLSG-----TYRELGAETGHPM 114
OY 81 -----TRFGSARA-----CN 91
DB 115 AEVFGCRGDLERAASLPAFCN 138

RESULT 11

T34656

hypoallergenic protein SC1A9.07 SC1A9.07 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T34656

R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream,

submitted to the EMBL data library, December 1998

A:Reference number: Z21552

A:Accession: T34656

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-251 <SAU>

A:Cross-references: EMBL:AL034446; PIDN:CAA22377.1; GSPDB:GN00070; SCOEDB:SC1A9.07

A:Experimental source: strain A3(2)

C:Genetics:
A:Gene: SCOEDB:SC1A9.07

Query Match 12.18: Score 66.5; DB 2; Length 251;
Best Local Similarity 33.3%; Pred. No. 12;
Matches 23; Conservative 4; Mismatches 29; Indels 13; Gaps 3;

OY 24 VNGSAFTWMSGPG-----CNNAERYSKGCSAIIHQKGYDESYTGRTAA-----LYN 71
DB 134 VNATFTDYMSSDGDVYETELAEVIRAGADGVFVPLTDPARIGSLAARFVPLNVLTY 193
OY 72 QACGSGVAH 80
DB 194 PAC-PGLAH 201

RESULT 12

DVMS2
multidrug resistance protein 2 mouse
N:Alternate names: P-glycoprotein MDR2
C:Species: Mus musculus (house mouse)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
R:Gros, P.; Raymond, M.; Bell, J.; Housman, D.
Mol. Cell. Biol. 8, 2770-2778, 1988
A:Title: Cloning and characterization of a second member of the mouse mdr gene family.
A:Reference number: A30409; MUID:88302195
A:Accession: A30409
A:Molecule type: mRNA
A:Residues: 11276 <RHS>

A:Cross-references: GB:J03398; NID:q199109; PIDN:AAA9516.1; PID:q387428
R:Kirschner, L.S.
Nucleic Acids Res. 24, 2824-2834, 1996
A:Title: De novo generation of simple sequence during gene amplification.
A:Reference number: S70711; MUID:96313253
A:Accession: S70711

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 43-92 <KIR>
A:Cross-references: EMBL:U46839; NID:q1228142; PIDN:ACG52722.1; PID:q1228143
C:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated to phospholipid antitumor drugs.
C:Genetics:

A:Gene: mdr2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane pr
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1049-1445/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:88_94/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted

Query Match 12.0%; Score 66; DB 1; Length 1276;
Best Local Similarity 20.4%; Pred. No. 62;
Matches 20; Conservative 19; Mismatches 31; Indels 28; Gaps 2;

OY 6 LEFSVTTVMILAMASEMYN-----GSAFTWMSGPCNNRERYSKGCCSAIHQKGYD 59
DB 333 VEFSLIGAFSVGOAPCIDAFANAGAAVIFDIIDNNPKIDSFSEHGKRDNIKNLE 392
OY 60 FS-----YTGTAAALYNOAGC 75
DB 393 FSDVHFSYPSRANIKILKGLNLYKSGGTVALVGNAGC 430

RESULT 13

T22396
hypothetical protein F49A5.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C:Accession: T22396
R:Mortimore, B.

Submitted to the EMBL Data Library, November 1996
A:Reference number: 219560
A:Accession: T22396

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-233 <MIL>

A:Cross-references: EMBL:Z81542; PIDN:CAB04418.1; GSPDB:GN00023; CESP:F49A5.6
A:Experimental source: clone F49A5
C:Genetics:

A:Gene: CESP:F49A5.6
A:Map position: 5
A:Introns: 86/3
C:Superfamily: thaumatin I

Query Match 11.9%; Score 65.5; DB 2; Length 233;
Best Local Similarity 26.0%; Pred. No. 15;
Matches 27; Conservative 16; Mismatches 42; Indels 19; Gaps 6;

OY 1 MASTKLFESVITVMILAMASE-----MYNGSAFTWMSG-PCNNRABRYSKGCSSAIHQ 54
DB 1 MALAQLIFAV-----SLALGAETRTITVYRCPFTIMPGIQGPNAG-----GGFTLHS 51
OY 55 KCGYDESYTGRTAA--LYNOACGSG--VAHTRFGSSARACNPF 94
DB 52 EHSRDTVSDSWTAGRIWARTGCDANFNCEGTGCGNSBQCNAG 95

RESULT 14

T04212
osmotin precursor - Arabidopsis thaliana
N:Alternate names: protein T5C23.80

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Aug-1999
C:Accession: T04212; S57524
R:Byran, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

Submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15261
A:Accession: T04212

A:Molecule type: DNA
A:Residues: 1-244 <BEV>
A:Cross-references: EMBL:AL049500

R:Capelli, N.; Simon, P.; Dionon, T.; Greppl, H.
submitted to the EMBL Data Library, June 1995
A:Description: Isolation of an osmotin cDNA clone from Arabidopsis.

A:Reference number: S57524
A:Accession: S57524
A:Molecule type: mRNA

A:Cross-references: EMBL:X89008; NID:9887389; PIDN:CAA61411.1; PID:9887390

A:Residues: 1-185, 'V', 187-244 <CAP>
C:Genetics:

A:Map position: 4
A:Introns: 132/2
A:Note: T5C23.80
C:Superfamily: thaumatin I
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-24/Product: osmotin #status predicted <MAT>

Query Match 11.9%; Score 65.5; DB 2; Length 244;
Best Local Similarity 23.1%; Pred. No. 15;
Matches 24; Conservative 19; Mismatches 44; Indels 17; Gaps 3;

OY 1 MASTKLFESVITVMILAMASEMYNGSAFTW--SGPCNNRABRYSKGCSSAIHQKGY 58
DB 1 MASTKLFESVITVMILAMASEMYNGSAFTW--SGPCNNRABRYSKGCSSAIHQKGY 58

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2002, 09:28:00 Search time 11.71 Seconds

(without alignments)
319.369 Million cell updates/sec

Title: US-09-882-434A-1

Perfect score: 551
Sequence: 1 MASTRLFFSVITVMMILIMAS.....FGSSARACNPFMKSFIFQC 102

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551	100.0	102	AMP1_MACIN	P80915 macadamia 1
2	75.5	13.7	485	HXB_YEAST	P04807 saccharomyc
3	71.5	13.0	403	PGK_CHLMU	O9pln4 chlamydia m
4	69	12.5	1281	MDR3_CRIGR	P23174 cricetus
5	68.5	12.4	1557	LMCL_CAEEL	O18823 caenorhabdi
6	68	12.3	1280	MDR1_HUMAN	P08183 homo sapien
7	67.5	12.3	902	VER_GVHA	P54232 heliothis a
8	66.5	12.1	151	CSGB_SALTY	P55226 salmoneila
9	66.5	12.1	157	RA05_ORYSA	Q01881 oryza sativ
10	66.5	12.1	162	RA17_ORYSA	Q01883 oryza sativ
11	66	12.0	1276	MDR2_MOUSE	P21440 mus musculu
12	65.5	11.9	244	OSL3_ARATH	P50700 arabidopsis
13	65	11.8	165	RA14_ORYSA	Q01882 oryza sativ
14	64.5	11.7	485	HXB_YEAST	P04806 saccharomyc
15	64	11.6	320	RLX1_STAAU	P12054 staphylococ
16	64	11.6	1278	MDR2_RAT	Q08201 rattus norv
17	64	11.6	1609	LMG1_HUMAN	P11047 homo sapien
18	63.5	11.5	166	RAG2_ORYSA	Q01885 oryza sativ
19	63.5	11.5	513	YAWL_RHISN	P55687 rhizobium s
20	63.5	11.5	828	MRC_KLEPN	P21647 klebsiella
21	63.5	11.5	1019	LFC_CANRO	Q26422 carcinoscor
22	63.5	11.5	1019	LFC_TACTR	P22975 tachyples
23	63	11.4	159	PRO4_LYCES	Q04108 lycopersico
24	63	11.4	1279	MDR3_HUMAN	P21439 homo sapien
25	62.5	11.3	235	THW2_THADA	P02884 thumathococ
26	62	11.3	118	CRYE_CRTPA	P52783 cryphonectr
27	62	11.3	159	PR06_LYCES	P04284 lycopersico
28	62	11.3	169	HSCC_DROME	P22979 dirosophila
29	61.5	11.2	455	YB12_YEAST	P38288 saccharomyc
30	61	11.1	501	AMT1_ARATH	P54144 arabidopsis
31	61	11.1	574	HMO3_SOLTU	O41438 solanum tub
32	61	11.1	1276	MDR1_CRIGR	P21448 cricetus
33	61	11.1	1786	LMB1_HUMAN	P07942 homo sapien

ALIGNMENTS

RESULT 1
ID AMP1_MACIN STANDARD: PRT: 102 AA.
AC P80915: Q04396;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ANTIMICROBIAL PEPTIDE 1 PRECURSOR (AMP1).
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota, Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-96.
RC TISSUE=Seed;
RX MEDLINE=97261828; PubMed=9108242;
RA Marcus J.P., Green J.L., Goulter K.C., Harrison S.J., Manners J.M.;
RT "Purification, characterisation and cDNA cloning of an antimicrobial
peptide from Macadamia integrifolia.";
RL Eur. J. Biochem. 244:743-749(1997).
CC -!- FUNCTION: ANTIMICROBIAL PEPTIDE WHICH INHIBITS THE GROWTH OF
A VARIETY OF FUNGI, OOMYCETES, GRAM-POSITIVE BACTERIAL
PHYTOPATHOGENS AND S.CEREVISIAE IN VITRO. NO ACTIVITY AGAINST
E.COLI.
CC -!- MISCELLANEOUS: ITS ANTIMICROBIAL ACTIVITY IS DIMINISHED BY CALCIUM
AND POTASSIUM CHLORIDE SALTS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL: Y10903; CAA71842.1; -!
CC Antibiologic; Fungicide; Signal.
CC FT SIGNAL 1 26
CC FT CHAIN 27 102
CC SO SEQUENCE 102 AA: 10943 MW: 9FB79954454B9311 CRC64;

Query Match 100.0%; Score 551; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.9e-51;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MASTRLFFSVITVMMILIMASFMVNGSAFYTWGSGCGNNRAERYSKCCCSA1HOKGIDF 60
DB 1 MASTRLFFSVITVMMILIMASFMVNGSAFYTWGSGCGNNRAERYSKCCCSA1HOKGIDF 60
OY 61 SYTGOTAAALYNAGCGGVAHTRFFGSSARACNPFMKSFIFQC 102
DB 61 SYTGOTAAALYNAGCGGVAHTRFFGSSARACNPFMKSFIFQC 102

RESULT 2

HXB_YEAST STANDARD: PRT: 485 AA.

ID HXB_YEAST

AC P04807:

DT 13-AUG-1987 (Rel. 05, Created).

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HEXOKINASE B (EC 2.7.1.1) (HEXOKINASE PII).

CN HKK2 OR HKB OR HEX1 OR YGL253W OR NRBA86.

OS Saccharomyces cerevisiae (Baker's yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

CC NCBI_TaxID=4932;

ON

RP SEQUENCE FROM N.A.

RP MEDLINE=86120382; PubMed=3003701;

RA Stacheler C., Stacheler J., Swan J., Botstein D., Konigsberg W.;

RT "Identification, cloning and sequence determination of the genes

RT specifying hexokinase A and B from yeast."

RL Nucleic Acids Res. 14:945-963(1986).

RN

RP SEQUENCE FROM N.A.

RP MEDLINE=8605643; PubMed=3905511;

RA Froehlich K.-U., Entian K.-D., Mecke D.;

RT "The primary structure of the yeast hexokinase PII gene (HKK2) which

RT is responsible for glucose repression."

RL Gene 36:105-111(1985).

RN

RP SEQUENCE FROM N.A.

RP STRAIN=SS288C / FY1679;

RC MEDLINE=97127827; PubMed=8972578;

RA Colisac E., Maillier E., Robineau S., Netter P.;

RT "Sequence of a 39,411 bp DNA fragment covering the left end of

RT chromosome VII of Saccharomyces cerevisiae."

RL Yeast 12:155-156(1996).

RN

RP SEQUENCE OF 1-246 FROM N.A.

RP STRAIN=W303;

RC MEDLINE=93311123; PubMed=8322518;

RA Breitwieser W., Price C., Schuster T.;

RT "Identification of a gene encoding a novel zinc finger protein in

RT Saccharomyces cerevisiae."

RL Yeast 9:551-556(1993).

RN

RP SEQUENCE OF 118-126; 175-184 AND 303-313.

RP STRAIN=ATCC 38531 / Y41;

RC MEDLINE=95255188; PubMed=7737086;

RA Norbeck J., Blomberg A.;

RT "Gene linkage of two-dimensional polyacrylamide gel electrophoresis

RT resolved proteins from isogenic families in Saccharomyces cerevisiae

RT by microsequencing of in-gel trypsin generated peptides."

RL Electrophoresis 16:149-156(1995).

RN

RP PHOSPHORYLATION OF SER-14.

RP MEDLINE=94114477; PubMed=8286332;

RA Kriegl T.M., Rush J., Vojtek A.B., Clifton D., Fraenkel D.G.;

RT "In vivo phosphorylation site of hexokinase 2 in Saccharomyces

RT cerevisiae."

RL Biochemistry 33:148-152(1994).

RN

RP PHOSPHORYLATION OF SER-157.

RP MEDLINE=97199316; PubMed=9047292;

RA Heidrich K., Otto A., Behlke J., Wenzel K.W., Kriegl T.;

RT "Autophosphorylation-inactivation site of hexokinase 2 in

RT Saccharomyces cerevisiae."

RL Biochemistry 36:1960-1964(1997).

RN

RP SEQUENCE OF 1-18, AND PHOSPHORYLATION OF SER-14.

RP MEDLINE=98384167; PubMed=9718324;

RA Behlke J., Heidrich K., Naumann M., Mueller E.-C., Otto A., Reuter R.,

RA Kriegl T.;

RT "Hexokinase 2 from Saccharomyces cerevisiae: regulation of oligomeric

RT structure by in vivo phosphorylation at serine-14."

RL Biochemistry 37:11989-11995(1998).

RN [9]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).

RX MEDLINE=78244654; PubMed=355643;

RA Anderson C.M., Stenkamp R.E., Steltz T.A.;

RT "Sequencing a protein by X-ray crystallography. II. Refinement of

RT yeast hexokinase B co-ordinates and sequence at 2.1-A resolution."

RL J. Mol. Biol. 123:15-33(1978).

CC -1- FUNCTION: MAIN GLUCOSE PHOSPHORYLATING ENZYME. MAY PLAY A

CC REGULATORY ROLE IN BOTH INDUCTION AND REPRESSION OF GENE

CC EXPRESSION BY GLUCOSE.

CC -1- CATALYTIC ACTIVITY: ATP + D-HEXOSE -> ADP + D-HEXOSE 6-PHOSPHATE.

CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE

CC INHIBITION BY ATP.

CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.

CC -1- SUBUNIT: HOMODIMER.

CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING

CC ISOENZYMES. DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.

CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.

CC -1- DATABASE: NAME=worthington-biochem.com/manual/H/K.htm

CC WWW="http://www.worthington-biochem.com/manual/H/K.htm".

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isdb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

DR EMBL: X03483; CAA27203.1; -

DR EMBL: M1181; AAA34697.1; -

DR EMBL: X94357; CAA64134.1; -

DR EMBL: 272775; CAA96973.1; -

DR EMBL: X67787; CAA48003.1; -

DR PIR: B23523; KIBYB.

DR PIR: S28555; S28555.

DR PDB: 2YHX; 1S-0UL-92.

DR SWISS-2DPAGE: P04807; YEAST.

DR YEPP: 8536; -

DR YEPP: 8548; -

DR SGD: S0003222; HKK2.

DR InterPro: IPR001312; Hexokinase.

DR Pfam: PF00349; hexokinase.1.

DR PRINTS: PR00475; HEXOKINASE.

DR ProDom: PD001109; Hexokinase.1.

DR PROSITE: PS00378; HEXOKINASES; 1.

KW Transferase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;

KW 3D-structure; Phosphorylation.

FT INIT-MET 0

FT BINDING 110

FT DOMAIN 151 177

FT MOD.RES 157 157

FT CONFLICT 28 28

FT CONFLICT 32 32

FT CONFLICT 60 60

FT CONFLICT 196 196

FT CONFLICT 420 421

FT CONFLICT 443 444

FT CONFLICT 452 452

FT CONFLICT 461 461

FT HELIX 23 33

FT HELIX 37 55

FT STRAND 65 66

FT STRAND 79 86

FT STRAND 90 99

FT TURN 100 101

FT STRAND 102 110

FT STRAND 112 112

FT TURN 115 118

FT TURN 123 123

FT HELIX 124 141

FT TURN 142 142

ATP (BY SIMILARITY).

GLUCOSE-BINDING (POTENTIAL).

PHOSPHORYLATION.

N -> I (IN REF. 1).

I -> N (IN REF. 2).

G -> V (IN REF. 1).

T -> S (IN REF. 1).

YN -> ST (IN REF. 2).

TS -> PH (IN REF. 2).

I -> V (IN REF. 2).

A -> P (IN REF. 2).


```

FT STRAND 150 155
FT STRAND 164 166
FT TURM 165 166
FT STRAND 169 176
FT TURM 175 176
FT STRAND 187 187
FT HELIX 188 199
FT STRAND 202 208
FT HELIX 210 221
FT TURM 223 224
FT STRAND 225 231
FT STRAND 235 241
FT HELIX 243 245
FT TURM 251 252
FT STRAND 263 266
FT TURM 270 275
FT HELIX 283 291
FT HELIX 299 304
FT HELIX 306 322
FT TURM 323 324
FT TURM 338 339
FT TURM 343 343
FT HELIX 344 351
FT HELIX 358 368
FT TURM 369 369
FT HELIX 374 395
FT TURM 396 396
FT HELIX 397 406
FT TURM 407 407
FT STRAND 411 416
FT TURM 418 421
FT TURM 424 425
FT HELIX 426 438
FT HELIX 445 447
FT STRAND 450 454
FT TURM 458 460
FT HELIX 461 475
SQ SEQUENCE 485 AA; 53811 MW; EAA7EC79612532AE CRC64;

```

Query Match 13.78% Score 75.5; DB 1; Length 485;
 Best Local Similarity 30.2%; Pred. No. 0.97;
 Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;

```

OY 8 FSVITVMM--IIAMSEWNGSAFTVWSPGCGNNAERYKSCGSAIHQKGYDPSYTG 64
DB 368 FGINTVQERKLRRLRSLTIG-----RAALSLVCGIAICQKRGYK---TG 411
OY 65 OFAALYNQAGSGVAHTRF-GSSARACNP-----FGW 95
DB 412 HIAA-----DGSVYNRRPGPREKAKANALAKDIYCW 440

```

```

RESULT 3
PGK_CHLMU STANDARD: PRT: 403 AA.
AC 09PLN4:
DT 20-AUG-2001 (Rel. 40; Created)
DT 20-AUG-2001 (Rel. 40; Last sequence update)
DE 20-AUG-2001 (Rel. 40; Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN PGK OR TC0065.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=MOPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Petersen J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Knoult-H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,

```

```

RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP + 3-
CC PHOSPHO-D-GLYCEROYL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE002274; AAF73528.1; -.
DR TIGR: TC0065;
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR001576; PGK.
DR Pfam: PF00162; PGK_1.
DR PRINTS: PR00477; PHGLYCKINASE.
DR PROSITE: PS00111; POLYGLYCERATE_KINASE.1.
KW Transferase; Kinase; GLYCOLYSIS; Complete proteome.
SQ SEQUENCE 403 AA; 43391 MW; 7A80C7A550B89F64 CRC64;

```

Query Match 13.08% Score 71.5; DB 1; Length 403;
 Best Local Similarity 29.3%; Pred. No. 2.2;
 Matches 24; Conservative 13; Mismatches 28; Indels 17; Gaps 4;

```

OY 17 IAMSEWNGSAFTVWSPGCGNNAERYKSCGSAIHQKGYDPSYTG 66
DB 307 IALFEVITQDSATVFWNGPVGEVPEPDGSKAIAOCLASHSSAVTVVG-----GDA 360
OY 67 AALYNQAGC-SGVAHTRFGSSA 87
DB 361 AAVALAGCTQISIVSTGGCA 382

```

```

RESULT 4
MDR3_CRIGR STANDARD: PRT: 1281 AA.
AC P23174;
DT 01-NOV-1991 (Rel. 20; Created)
DT 01-NOV-1991 (Rel. 20; Last sequence update)
DT 01-NOV-1997 (Rel. 35; Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPROTEIN 3).
GN PGY3 OR PGP3.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=92135896; PubMed=1685679;
RA Endicott J.A., Sarangi F., Ling V.;
RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
RT gene family."
RL DNA Seq. 2:89-101(1991).
CC -1- FUNCTION: ENERGY-INDEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CC CANNOT.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MDR SUBFAMILY.
CC

```

CC	- I - SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC	-----
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb.ch/announce)
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U46673; AAC48152.1; -
DR	HSSP; P02468; 1TPE.
DR	Wormpep; C54D1.5; CE06981.
DR	InterPro; IPRO000561; EGF-like.
DR	InterPro; IPRO01886; LamNT.
DR	InterPro; IPRO00034; Laminin_B.
DR	InterPro; IPRO02049; Laminin_EGF.
DR	Pfam; PF00053; Laminin_B; 1.
DR	Pfam; PF00052; Laminin_EGf; 10.
DR	Pfam; PF00055; Laminin_Nterm; 1.
DR	PRINTS; PR00011; EGFLAMININ.
DR	ProDom; PD002082; LamNT; 1.
DR	ProDom; PD003031; Laminin_B; 1.
DR	SMART; SM00180; EGF_Lam; 9.
DR	SMART; SM00001; EGF_Like; 2.
DR	SMART; SM00281; Lamb; 1.
DR	SMART; SM00136; LamNT; 1.
DR	PROSITE; PS00022; EGF_1; 8.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
KW	Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.
FT	SIGNAL 1 ?
FT	CHAIN ? 1557
FT	DOMAIN ? 271
FT	LAMININ N-TERMINAL (DOMAIN VI).
FT	DOMAIN 272 331
FT	LAMININ EGF-LIKE 1.
FT	DOMAIN 332 387
FT	LAMININ EGF-LIKE 2.
FT	DOMAIN 388 434
FT	LAMININ EGF-LIKE 3.
FT	DOMAIN 435 487
FT	LAMININ EGF-LIKE 4.
FT	DOMAIN 488 497
FT	LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	DOMAIN 498 688
FT	LAMININ DOMAIN IV.
FT	DOMAIN 687 732
FT	LAMININ EGF-LIKE 6 (C-TERMINAL).
FT	DOMAIN 737 769
FT	LAMININ EGF-LIKE 7 (INCOMPLETE).
FT	DOMAIN 770 819
FT	LAMININ EGF-LIKE 8.
FT	DOMAIN 820 874
FT	LAMININ EGF-LIKE 9.
FT	DOMAIN 875 930
FT	LAMININ EGF-LIKE 10.
FT	DOMAIN 931 978
FT	LAMININ EGF-LIKE 11.
FT	DOMAIN 979 1025
FT	DISULFID 272 281
FT	DISULFID 274 295
FT	DISULFID 297 306
FT	DISULFID 297 306
FT	DISULFID 309 329
FT	DISULFID 309 329
FT	DISULFID 332 341
FT	DISULFID 332 341
FT	DISULFID 334 357
FT	DISULFID 334 357
FT	DISULFID 360 369
FT	DISULFID 360 369
FT	DISULFID 372 385
FT	DISULFID 372 385
FT	DISULFID 388 400
FT	DISULFID 388 400
FT	DISULFID 390 406
FT	DISULFID 390 406
FT	DISULFID 408 417
FT	DISULFID 408 417
FT	DISULFID 420 432
FT	DISULFID 420 432
FT	DISULFID 435 449
FT	DISULFID 435 449
FT	DISULFID 437 456
FT	DISULFID 437 456
FT	DISULFID 458 467
FT	DISULFID 458 467
FT	DISULFID 470 485
FT	DISULFID 470 485
FT	DISULFID 770 778
FT	DISULFID 770 778
FT	DISULFID 772 788
FT	DISULFID 772 788
FT	DISULFID 791 800
FT	DISULFID 791 800
FT	DISULFID 803 817
FT	DISULFID 803 817
FT	DISULFID 820 834
FT	DISULFID 820 834
FT	DISULFID 822 841
FT	DISULFID 822 841
FT	DISULFID 844 853
FT	DISULFID 844 853
FT	DISULFID 856 872
FT	DISULFID 856 872
FT	DISULFID 875 894
FT	DISULFID 875 894
FT	DISULFID 877 901
FT	DISULFID 877 901

```

FT DISULFID 903 912 BY SIMILARITY.
FT DISULFID 915 928 BY SIMILARITY.
FT DISULFID 931 943 BY SIMILARITY.
FT DISULFID 933 950 BY SIMILARITY.
FT DISULFID 952 961 BY SIMILARITY.
FT DISULFID 964 976 BY SIMILARITY.
FT DISULFID 979 991 BY SIMILARITY.
FT DISULFID 981 998 BY SIMILARITY.
FT DISULFID 1000 1009 BY SIMILARITY.
FT DISULFID 1012 1023 BY SIMILARITY.
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1168 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1557 AA; 172723 MW; CAF0B51F8D5E8D2F CRC64;

```

```

Query Match 12.4% Score 68.5; DB 1; Length 1557;
Best Local Similarity 28.8%; Pred. No. 15;
Matches 19; Conservative 5; Mismatches 33; Indels 9; Gaps 2;

```

```

OY 29 FTWSPGCGNNAERYSKCCSAIHOKGYDFSYGTOTALYNAQ--CSGVNHRFGSS 86
DB 920 YNTSGIGC-----QECNCPLGSEGNVCVNTGCGCCCKPVTGRCRDCAHYHFGS 972
OY 87 ARACNP 92
DB 973 ANGCOP 978

```

```

RESULT 6
ID MDRL_HUMAN STANDARD; PRT; 1280 AA.
AC P08183; Q12755; Q14812;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1).
GN ABCB1 OR PGY1 OR MDRL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=67028230; PubMed=2876781;
RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
RT "Intenal duplication and homology with bacterial transport proteins
RT in the mdrl (P-glycoprotein) gene from multidrug-resistant human
RT cells";
RL Cell 47:381-389(1986);
RN [2]
RX MEDLINE=90094448; PubMed=1967175;
RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
RT "Genomic organization of the human multidrug resistance (MDR1) gene
RT and origin of P-glycoproteins";
RL J. Biol. Chem. 265:506-514(1990);
RN [3]
RX MEDLINE=97190336; PubMed=9038218;
RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffezeou J.P.,

```

```

RA Dumontet C., Slik B.I.;
RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
RT altered phenotype, and resistance to cyclosporins.";
RL J. Biol. Chem. 272:5974-5982(1997).
RN [4]
RX MEDLINE=90290529; PubMed=1972623;
RA Smith A., Beck C., Gibson A.;
RT Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RX MEDLINE=90290529; PubMed=1972623;
RA Gekeler V., Weger S., Probst H.;
RT "mdrl/P-glycoprotein gene segments analyzed from various human
RT leukemic cell lines exhibiting different multidrug resistance
RT profiles";
RL Biochem. Biophys. Res. Commun. 169:796-802(1990).
RN [6]
RX MEDLINE=90290529; PubMed=1972623;
RA Kioka N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,
RA Pastan I., Ueda K.;
RT Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE
CC AND BRAIN.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M14758; AAA59575.1; -
DR EMBL: M29447; AAA59576.1; JOINED.
DR EMBL: M29424; AAA59576.1; JOINED.
DR EMBL: M29425; AAA59576.1; JOINED.
DR EMBL: M29426; AAA59576.1; JOINED.
DR EMBL: M29427; AAA59576.1; JOINED.
DR EMBL: M29428; AAA59576.1; JOINED.
DR EMBL: M29429; AAA59576.1; JOINED.
DR EMBL: M29430; AAA59576.1; JOINED.
DR EMBL: M29431; AAA59576.1; JOINED.
DR EMBL: M29432; AAA59576.1; JOINED.
DR EMBL: M29433; AAA59576.1; JOINED.
DR EMBL: M29434; AAA59576.1; JOINED.
DR EMBL: M29435; AAA59576.1; JOINED.
DR EMBL: M29436; AAA59576.1; JOINED.
DR EMBL: M29437; AAA59576.1; JOINED.
DR EMBL: M29438; AAA59576.1; JOINED.
DR EMBL: M29439; AAA59576.1; JOINED.
DR EMBL: M29440; AAA59576.1; JOINED.
DR EMBL: M29441; AAA59576.1; JOINED.
DR EMBL: M29442; AAA59576.1; JOINED.
DR EMBL: M29443; AAA59576.1; JOINED.
DR EMBL: M29444; AAA59576.1; JOINED.
DR EMBL: M29445; AAA59576.1; JOINED.
DR EMBL: M29446; AAA59576.1; JOINED.
DR EMBL: AF016535; AAB69423.1; -
DR EMBL: AC002457; AAC82531.1; -
DR EMBL: M37724; AAA88047.1; -
DR EMBL: M37725; AAA88048.1; -
DR EMBL: X58723; CAA11558.1; -
DR PIR: A25059; DVH01.
DR PIR: A34914; A34914.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR001140; ABC_transport_tmnm.

```

DR InterPro: IPR001687; ATP-GTP-A.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR SMART: SM00382; AAA_2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 DR ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 MW MultiGene Family.
 FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 216 236 POTENTIAL.
 FT TRANSMEM 297 317 POTENTIAL.
 FT TRANSMEM 326 346 POTENTIAL.
 FT DOMAIN 347 710 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 711 731 POTENTIAL.
 FT TRANSMEM 757 777 POTENTIAL.
 FT TRANSMEM 833 853 POTENTIAL.
 FT TRANSMEM 854 874 POTENTIAL.
 FT TRANSMEM 937 957 POTENTIAL.
 FT TRANSMEM 974 994 POTENTIAL.
 FT DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 427 434 ATP (BY SIMILARITY).
 FT NP_BIND 1070 1077 ATP (BY SIMILARITY).
 FT REPEAT 1 637
 FT CARBOHYD 638 1280
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 23 23 S -> R (IN REF. 6).
 FT CONFLICT 185 185 G -> V (IN REF. 1 AND 3).
 FT CONFLICT 336 336 MISSING (IN REF. 3).
 FT CONFLICT 412 412 G -> A (IN REF. 3).
 FT CONFLICT 438 438 O -> S (IN REF. 3).
 SQ SEQUENCE 1280 AA; 141462 MW; ABIC279531F43675 CRC64;

Query Match 12.3%; Score 68; DB 1; Length 1280;
 Best Local Similarity 24.5%; Pred. No. 14;
 Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LEFSVITVMDLMASEV-----NSAFTVSGPCNNRAEYKSCGSAIHQKGYD 59
 DB 334 VFSVILGANSVGASPSIEAFANAKAAEIKTIDNKPSIDSYSGHKPKINIKGLE 393
 QY 60 -----FSY-----TSGTAAALYNQAC 75
 DB 394 FRNVHFSYPSRKEVKILMLNLKAYQSGTVALDGNSSC 431
 RESULT 7
 ID VEF_GVHA STANDARD; PRT; 902 AA.
 AC P54232;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VIRAL ENHANCING FACTOR (VEF) (ENHANCIN) (104 KDA GLYCOPROTEIN)
 DE (SYNERGISTIC FACTOR).
 GN VEF.
 OS Heliothis armigera granulosis virus (hagv) (Heliothis armigera
 OS granulovirus).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
 OX NCBI_TaxID=45440;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96068802; PubMed=7595376;
 RA Reelink P.W., Corsaro B.G., Granados R.R.;
 RT "Characterization of the Helicoverpa armigera and Pseudaletia
 RT unipuncta granulovirus enhancer genes.";
 RL J. Gen. Virol. 76:2693-2705(1995).
 CC -1- FUNCTION: INVOLVED IN DISRUPTION OF THE PERTROPHIC MEMBRANE AND
 CC FUSION OF NUCLEOCAPSIDS WITH MIDGUT CELLS (BY SIMILARITY).

CC -1- SIMILARITY: TO TNGV AND PUGV VEF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D28558; BAA05908.1; -
 KW Glycoprotein; Late protein.
 FT CARBOHYD 73 73
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 840 840 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 902 AA; 104791 MW; DB45BAA5F675FDE CRC64;

Query Match 12.3%; Score 67.5; DB 1; Length 902;
 Best Local Similarity 28.6%; Pred. No. 12;
 Matches 20; Conservative 10; Mismatches 33; Indels 7; Gaps 4;

QY 26 GSAF-TVWSGPCNNRAE-RYSKCGSAIHQK-GYDFSYTGOT--AALYNQACGCV 78
 DB 213 GCAYGAFMTAPASTNLGELRVSPTNMVMVHELHGAQDFVTVMRLIEINNSFCORI 272
 QY 79 AHTRGSSAR 88
 DB 273 QYTMNKTTR 282

RESULT 8
 ID CSGB_SALTY STANDARD; PRT; 151 AA.
 AC P55226;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MINOR CURLIN SUBUNIT PRECURSOR (FIMBRIN SEPT7 MINOR SUBUNIT).
 GN CSGB OR AGFB.
 OS Salmonella typhimurium, and Salmonella enteritidis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OX NCBI_TaxID=602, 592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhimurium; STRAIN-SR-11;
 RA MEDLINE=98117058; PubMed=9457880;
 RX Rowling U., Bian Z., Hammar M., Sieralta W.D., Normark S.;
 RT "Curli fibers are highly conserved between Salmonella typhimurium and
 RT Escherichia coli with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. enteritidis; STRAIN-27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBROBLASTS. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: STRONG, TO E. COLI CSGB.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ002301; CAA05316.1; -
DR EMBL: U43280; AAC43598.1; -
DR StyGene: SGI0609; cs9d.
KW EMBL: Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SO SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;

Query Match
Best Local Similarity 12.1%; Score 66.5; DB 1; Length 151;
Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;

QY 30 TVWSGPCNNRAREYKSCGSAIHQKGYDFSTGTALYNGCGSVAHRTFGSSA 87
DB 69 SVISQEGGNRRK-----VDQAGNYNFAYTEOT---GNANDASISQSAVGNSSA 113

RESULT 9
RA05_ORYSA STANDARD: PRT; 157 AA.
AC 001861;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE SEED ALLERGENIC PROTEIN RA5 PRECURSOR.
GN RA5.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN 11;
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=93144699; PubMed=7678765;
RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,
RA Nakamura R., Matsuda T.;
RT "Gene structure and expression of rice seed allergenic proteins
RT belonging to the alpha-amylase/trypsin inhibitor family.";
RT Plant Mol. Biol. 21:239-248(1993).
RL -1- PTM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D11430; BAA01996.1; -
DR PIR: S31078; S31078.
DR HSSP: P01085; 1HSS.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals_tryp_aml_1nh.
DR Pfam: PF00234; tryp_alpha_aml_1.
DR PRINTS: PRO0808; AMLASEINHTR.
DR PRINTS: PRO0809; RAGALLERGEN.
DR SMART: SM00499; AAI.
DR PROSITE: PS00426; CEREAL_TRYP_AMYL_1NH.
DR Allergen: Multigene family; Signal.

```

```

FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 157 SEED ALLERGENIC PROTEIN RA5.
SO SEQUENCE 157 AA; 17118 MW; C8A5495F8FB399E6 CRC64;

Query Match
Best Local Similarity 12.1%; Score 66.5; DB 1; Length 157;
Matches 33; Conservative 12; Mismatches 35; Indels 69; Gaps 7;

QY 1 MASTKLFEVS--ITYMMLTAMASEVN-----GSAFTVWSGPCNNRAE 42
DB 1 MASNKVFESVLLAVSVLAATATMAEYHODQVYTRARCPGCMGYPMYSLPRCALVK 60

QY 43 RX-----SKCGSAI-HQKGYDFSTGTALYNGCGSG 77
DB 61 RQCRGSAAMAEVRRDCCQOLAVDSDWRCRAISHMLGG-----TYRELGAPD 109

QY 78 VAH-----TRFGSSARA-----CN 91
DB 110 VGHPMSEVRGCRGDDLERAASLPAPCN 138

RESULT 10
RA17_ORYSA STANDARD: PRT; 162 AA.
AC 001883;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE SEED ALLERGENIC PROTEIN RA17 PRECURSOR.
GN RA17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN 11;
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=93144699; PubMed=7678765;
RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,
RA Nakamura R., Matsuda T.;
RT "Gene structure and expression of rice seed allergenic proteins
RT belonging to the alpha-amylase/trypsin inhibitor family.";
RT Plant Mol. Biol. 21:239-248(1993).
RL [2]
RN 12;
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=92289999; PubMed=1376283;
RA Izumi H., Adachi T., Fujii N., Matsuda T., Nakamura R., Tanaka K.,
RA Urisu A., Kurosawa Y.;
RT "Nucleotide sequence of a cDNA clone encoding a major allergenic
RT protein in rice seeds. Homology of the deduced amino acid sequence
RT with members of alpha-amylase/trypsin inhibitor family.";
RL FEBS Lett. 302:213-216(1992).
CC -1- PTM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X66257; CAA46983.1; -
DR EMBL: D11431; BAA01997.1; -
DR PIR: S21157; S21157.
DR HSSP: P01085; 1HSS.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals_tryp_aml_1nh.

```

```

DR HSSP: P13569; INBD.
DR MGD: MGI:97569; PGY2.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001140; ABC_transpoftr_tmem.
DR InterPro: IPR001687; ATP_gtp_A.
DR Pfam: PF00664; ABC_membrane_2.
DR Pfam: PF00005; ABC_tran: 2.
DR SMART: SM00382; AAA: 2.
DR PROSITE: PS00211; ABC_TRANSPORTER: 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
KW DOMAIN 1.
FT TRANSMEM 52 75 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 346 707 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 708 728 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 849 POTENTIAL.
FT TRANSMEM 850 870 POTENTIAL.
FT TRANSMEM 933 953 POTENTIAL.
FT TRANSMEM 970 990 POTENTIAL.
FT DOMAIN 991 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 635
FT REPEAT 636 1276
SQ SEQUENCE 1276 AA; 140332 MW; A6C38DCD2C1BEEF CRC64;

Query Match 12.0%; Score 66; DB 1; Length 1276;
Best Local Similarity 20.4%; Pred. No. 22;
Matches 20; Conservative 19; Mismatches 31; Indels 28; Gaps 2;

OY 6 LEFSYIVVMMLIAMSSEVNVN-----GSAFTVWSGPGCNRRARERKSKCCSAIHQKGYD 59
      :||::: : : : : :||::: : : : : :||:
Db 333 VFFSILICAFVSGQAPCIDAFANARGAAVYFDIIDNNPKIDSPFSERGHKRPDKINMLE 392
OY 60 FS-----YTGQTALYNQAG 75
      || : : : :||
Db 393 FSDVHFSTPSRANIKIKLGNLKVSSGQTVALVGNSGC 430

RESULT 12
OSL3_ARATH
ID OSL3_ARATH STANDARD; PRT; 244 AA.
AC P50700.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OSMOTIN-LIKE PROTEIN OSM34 PRECURSOR.
GN OSM34.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV, COLUMBIA; TISSUE=leaf;
RC MEDLINE=97354294; PubMed=9210588;
RA Capelli N., Dlogon T., Greppin H., Simon P.;
RT "Isolation and characterization of a cDNA clone encoding an
RL osmotin-like protein from Arabidopsis thaliana.";
RL Gene 191:51-56(1997).
CC -I- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X89008; CAA61411.1; .
 DR HSSP: P25871; IAUIN.
 DR InterPro: IPR001938; Thaumatin.
 DR Pfam: PF00314; Thaumatin; 1.
 DR PRINTS: PR00347; THAUMATIN.
 DR ProDom: PD001321; Thaumatin; 1.
 DR SMART: SM00205; THN; 1.
 DR PROSITE: PS00316; THAUMATIN; 1.
 KM SIGNAL.
 FT CHAIN 1 22 POTENTIAL.
 FT DISULFID 23 244 OSMOTIN-LIKE PROTEIN OSM34.
 FT DISULFID 31 222 BY SIMILARITY.
 FT DISULFID 72 82 BY SIMILARITY.
 FT DISULFID 87 93 BY SIMILARITY.
 FT DISULFID 138 212 BY SIMILARITY.
 FT DISULFID 143 195 BY SIMILARITY.
 FT DISULFID 151 161 BY SIMILARITY.
 FT DISULFID 165 174 BY SIMILARITY.
 FT DISULFID 175 182 BY SIMILARITY.
 SQ SEQUENCE 244 AA; 26603 MW; 9FBE8956E9F286E0 CRC64;

Query Match 11.9%; Score 65.5; DB 1; Length 244;

Best Local Similarity 23.1%; Pred. No. 6;

Matches 24; Conservative 19; Mismatches 44; Indels 17; Gaps 3;

OY 1 MASTKLFPSVTVVMKLIAMSEVNGSAFTW--SGFGCNRRERYSKCCSIAHOKGY 58
 DB 5 LVSFFIFALLISVAVTAAPETLNGCSYVWMAASPGGGRID-----AGQSWRL 55
 OY 59 DFTYTGOTALLYNOAGCGVAHTRFGSSARACNPFQKKSIFIOC 102
 DB 56 DVAAGTKMARIMGRITCN-----FDSSGRCRCOTGDCSGGLQC 93

RESULT 13
 ID RA14_ORYSA STANDARD; PRT: 165 AA.
 AC Q01882;

DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE SEED ALLERGENIC PROTEIN RA14 PRECURSOR.

GN RA14.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Seed;

RX MEDLINE=93144699; PubMed=7678765;

RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,

RA Nakamura R., Matsuda S.,

RT "Gene structure and expression of rice seed allergenic proteins
 belonging to the alpha-amylase/trypsin inhibitor family.";

RL Plant Mol. Biol. 21:239-248(1993).

CC -1- PFM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CEREAL TRIPSIN/ALPHA-AMYLASE INHIBITOR
 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D11432; BAA01998.1; .
 DR PIR: S31080; S31080.
 DR HSSP: P01085; IHS8.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal tryp. amyl. inh.
 DR Pfam: PF00234; tryp_alpha_amyl; 1.
 DR PRINTS: PR00808; AMYLASEINHTR.
 DR PRINTS: PR00809; RAGALLERGEN.
 DR SMART: SM00499; AAI; 1.
 DR PROSITE: PS00426; CEREAL_TRYP_AMYL_INH; 1.
 KM Allergen; Multigene family; signal.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 165 SEED ALLERGENIC PROTEIN RA14.
 FT DISULFID 165 178 BY SIMILARITY.
 SQ SEQUENCE 165 AA; 17887 MW; 13F99783726CEFC6A CRC64;

Query Match 11.8%; Score 65; DB 1; Length 165;

Best Local Similarity 20.4%; Pred. No. 4.8;

Matches 31; Conservative 17; Mismatches 32; Indels 72; Gaps 7;

OY 1 MASTKLFPS--VITVMKLIAMSEVNG-----GSAFTWGGGCGNNRAE 42
 DB 1 MASNKVFSAALLITIVSVLAATTRMADHNRDQVYSLGRCQPGKGYPTSLRCRAYVK 60
 OY 43 RY-----SKCGSAI-HQKGGYDFSYGOTALLYNOAG 74
 DB 61 RQCVGTSPGAVDQLADDCRELAAYVDSWCKGCSALNHVVG-----IYRELG 109
 OY 75 CSGVAH-----TRFGSSARA-----CN 91
 DB 110 AFDVGHMAVEVPCGRCRGLDERAASLPAFCN 141

RESULT 14
 ID HXKA_YEAST STANDARD; PRT: 485 AA.
 AC P04806;

DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HEXOKINASE A (EC 2.7.1.1) (HEXOKINASE PI).

GN HXK1 OR HKA OR YFR053C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86120382; PubMed=3003701;

RA Stachelek C., Stachelek J., Swan J., Botstein D., Konigsberg W.;

RT "Identification, cloning and sequence determination of the genes
 specifying hexokinase A and B from yeast.";

RL Nucleic Acids Res. 14:945-963(1986).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=86083199; PubMed=3908224;

RA Kopeckzi E., Entian K.-D., Mecke D.;

RT "Complete nucleotide sequence of the hexokinase PI gene (HXK1) of
 Saccharomyces cerevisiae.";

RL Gene 39:95-102(1985).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=5288C / AB972;

RX MEDLINE=95400292; PubMed=7670463;

RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,

RA Sasagawa S.-I., Sasagawa M., Tsuchiya Y., Soeda E., Yokoyama K.,

RA Yamazaki M., Tashiro H., Eki T.;

RT "Analysis of the nucleotide sequence of chromosome VI from
 Saccharomyces cerevisiae.";

RL Nat. Genet. 10:261-268(1995).

(4)
 RN SEQUENCE FROM N.A.
 RP STRAIN-S288C / AB972;
 RC MEDLINE-96287652; PubMed-8666379;
 RA Eki T., Naitou M., Hagihara H., Ozawa M., Sasagawa S.-I.,
 RA Sasagawa M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
 RT "Analysis of a 36.2 kb DNA sequence including the right telomere of
 RL chromosome VI from *Saccharomyces cerevisiae*.";
 RN Yeast 12:149-167(1996).
 [5]
 RP ATP-BINDING AND SEQUENCE OF 104-112.
 RX MEDLINE-86227998; PubMed-3131329;
 RA Tamura J.K., Ladine J.R., Cross R.L.;
 RT "The adenine nucleotide binding site on yeast hexokinase PII.
 RT Affinity labeling of Lys-111 by pyridoxal
 RL 5'-diphospho-5'-adenosine";
 RN J. Biol. Chem. 263:7907-7912(1988).
 [6]
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
 RX MEDLINE-81049624; PubMed-7001031;
 RA Bennett W.S., Jr., Steitz T.A.;
 RT "Structure of a complex between yeast hexokinase A and glucose. I.
 RT Structure determination and refinement at 3.5-A resolution.";
 RL J. Mol. Biol. 140:183-210(1980).
 CC -1- CATALYTIC ACTIVITY: ATP + D-HEXOSE = ADP + D-HEXOSE 6-PHOSPHATE.
 CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
 CC INHIBITION BY ATP.
 CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
 CC ISOENZYMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
 CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
 CC -1- DATABASE: NAME-WORTHINGTON-Biochem.com/manual/H/HK.html".
 CC WWW="http://www.worthington-biochem.com/manual/H/HK.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D44597; BAA08019.1;
 DR EMBL: M14410; AAA34698.1;
 DR EMBL: X03482; CAA27202.1;
 DR EMBL: D50617; BAA09292.1;
 DR PIR: A24531; KIRBHA.
 DR PIR: A28178; A28178.
 DR PDB: 1HKG; 15-OCT-91.
 DR SWISS-2DPAGE: P04806; YEAS
 DR YEPD: 7515;
 DR YEPD: 8524;
 DR SGD: S0001949; HXK1.
 DR InterPro: IPR001312; Hexokinase.
 DR Pfam: PF00349; hexokinase; 1.
 DR PRINTS: PR00475; HEXOKINASE.
 DR PRODOM: PD001109; Hexokinase; 1.
 DR PROSITE: PS00378; HEXOKINASES; 1.
 KW Transferrase; kinase; Glycolysis; Allosteric enzyme; ATP-binding;
 KW 3D-structure.
 FT BINDING 111
 FT DOMAIN 152 178
 FT CONFLICT 61
 FT CONFLICT 103
 FT CONFLICT 194
 FT CONFLICT 244
 FT CONFLICT 356
 FT CONFLICT 364
 FT CONFLICT 388
 FT CONFLICT 444
 FT CONFLICT 479
 FT TURN 19
 AP.
 GLUCOSE-BINDING (POTENTIAL).
 G -> V (IN REF. 1).
 H -> R (IN REF. 1).
 N -> K (IN REF. 1).
 V -> C (IN REF. 1).
 EN -> VF (IN REF. 2).
 I -> M (IN REF. 1).
 I -> T (IN REF. 2).
 D -> EN (IN REF. 1).
 SL -> VS (IN REF. 1).

FT	HELIX	21	29
FT	HELIX	32	33
FT	HELIX	38	44
FT	TURN	45	46
FT	HELIX	47	54
FT	TURN	55	56
FT	STRAND	66	67
FT	STRAND	82	82
FT	STRAND	85	85
FT	STRAND	88	88
FT	TURN	89	90
FT	STRAND	91	92
FT	STRAND	94	94
FT	STRAND	97	97
FT	STRAND	111	111
FT	STRAND	115	119
FT	TURN	121	122
FT	TURN	124	126
FT	TURN	128	130
FT	HELIX	131	140
FT	HELIX	142	143
FT	STRAND	161	161
FT	STRAND	171	171
FT	TURN	176	177
FT	STRAND	180	181
FT	TURN	189	195
FT	HELIX	196	200
FT	TURN	211	216
FT	HELIX	218	222
FT	TURN	224	225
FT	STRAND	226	232
FT	STRAND	236	239
FT	STRAND	242	242
FT	TURN	244	245
FT	STRAND	264	264
FT	STRAND	266	267
FT	HELIX	270	272
FT	TURN	286	287
FT	HELIX	288	292
FT	TURN	293	294
FT	HELIX	300	303
FT	TURN	304	305
FT	TURN	308	310
FT	HELIX	311	323
FT	TURN	324	325
FT	TURN	339	340
FT	HELIX	345	351
FT	TURN	359	369
FT	HELIX	375	377
FT	HELIX	378	396
FT	TURN	398	401
FT	HELIX	402	406
FT	TURN	407	408
FT	STRAND	412	417
FT	TURN	419	420
FT	HELIX	428	438
FT	STRAND	450	450
FT	STRAND	454	454
FT	TURN	458	460
FT	HELIX	461	474
FT	TURN	475	477
SO	SEQUENCE	485 AA;	53738 MW; AF5C9DA8F17BC3D0 CRC64;

Query Match 11.7%; Score 64.5; DB 1; Length 485;
 Best Local Similarity 27.4%; Pted. No. 14;
 Matches 17; Conservative 9; Mismatches 19; Indels 17; Gaps 2;
 QY 40 RAERYSKGCSATHKGYDFSYTGOTALVN-----QAGCGVAHTRFGSSARACPF 93
 Db 391 RAARLAVGIAIAICQKRGYKTHIAADSGVYKPKYGFKEAAGKLR-----DIY 439
 QY 94 GW 95

Db 440 GW 441

RESULT 15

ID RLX1_STAAU STANDARD; PRT; 320 AA.

AC P12054;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 01-OCT-1989 (Rel. 12, Last annotation update)

DE RLX PROTEIN.

OS Staphylococcus aureus.

OG Plasmid PS194.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/staphylococcus group; Staphylococcus.

OX NCBI_TaxID=1280;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88189810; PubMed=3357770;

RA Projan S.J., Moghazeh S., Novick R.P.;

RT "Nucleotide sequence of PS194, a streptomycin-resistance plasmid from

Staphylococcus aureus.";

RL Nucleic Acids Res. 16:2179-2187(1988).

CC -!- FUNCTION: THIS PROTEIN IS PROBABLY REQUIRED FOR RELAXATION

COMPLEX FORMATION AND PLASMID MOBILIZATION BY CONJUGATIVE

PLASMIDS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: X06627; CAA29840.1;

DR PIR: S00935; S00935.

KW Plasmid.

SO SEQUENCE 320 AA 36997 MW: 256691BFEE6D6FC4 CRC64;

Query Match 11.68; Score 64; DB 1; Length 320;

Best Local Similarity 30.78; Pred. No. 11;

Matches 31; Conservative 9; Mismatches 29; Indels 32; Gaps 6;

QY 1 MASTKLFESVITVWMLIAMASEVNGSAFTVWSGPCNNRAERYSKGCSATHQKGYDF 60

DB 1 MATTKLGNT-----KASRAINVA-----ERRAEKSGLNCVDVIK----- 37

QY 61 SYTGCTAALYNAGCSGV-AHTRFSG-----SARACNPG 94

DB 38 SYFKOTRALYKKE--NGVOAHTVIOFKEKGEVYAKKCEIG 76

Search completed: January 24, 2002, 09:29:50
Job time: 110 sec

